الاختبارات والمقاييس

تأليف

ليونا أ. تايلر

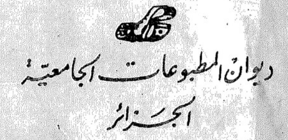
ترجمة

الدكتور سعد عبد الرحمن

مراجعة

الدكتور محمد عثمان نجاتي





ديوان المطبوعات الجامعية

الجزائر

الإختبارات
والمقاييس

أعيد طبع هذا الكتاب على مطابع ديوان المطبوعات الجامعية
بترخيص من الناشر الاصلي

بيروت : ص.ب: ٨٠٦٤ - هاتف: ٣١٥٨٥٩ - ٣١٥١٠١ - برقياً: داشروق - تلكس: SHOROK 20175 LE
القاهرة: ١٦ شارع جواد حسني - هاتف: ٧٧٤٨١٤ - ٧٧٤٥٧٨ - برقياً: شروق - تلكس: 93091 SHROK UN
SHOROUK INTERNATIONAL: 316/318 REGENT STREET, LONDON W1, UK, TEL: 637 2743/4, TELEX SHOROK 257786

بسم الله الرحمن الرحيم

مكتبة أصول علم النفس الحديث
بإشراف الدكتور محمد عثمان نجاتي

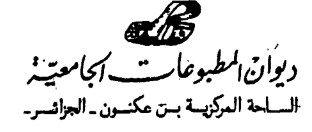
الاختبارات والمقاييس

تأليف
ليونا أ. تايلر

ترجمة
الدكتور سعد عبد الرحمن
استاذ علم النفس بجامعة الكويت

مراجعة
الدكتور محمد عثمان نجاتي
استاذ علم النفس بجامعة القاهرة
وجامعة الكويت





ديوان المطبوعات الجامعية
الساحة المركزية بن عكنون - الجزائر -

هذا الكتاب ترجمة لكتاب

Leona E. Tyler: **Tests and Measurements**

2nd edition. Englewood Cliffs, New Jersey: Prentice-Hall, Inc, 1971.

مكتبة أصول علم النفس الحديث

إن النمو الهائل لعلم النفس والتحامه المتزايد مع العلوم الإجتماعية والبيولوجية قد جعل من الضروري البحث عن أساليب جديدة لتعليمه في المستويات الأولى من التعليم الجامعي. ولم نعد بعد نشعر بالرضا عن المقرر الأساسي التقليدي الذي يُكَيّف عادة لكتاب واحد يحاول أن يعرض كل شيء عرضاً خفيفاً يضحي بالعمق في سبيل الشمول. إن علم النفس قد أصبح متعدد النواحي بحيث لم يعد من المتيسر لأي شخص واحد، أو عدد قليل من الأشخاص، أن يكتبوا فيه عن تمكن تام. والبديل الآخر لذلك وهو الكتاب الذي يهمل ميادين كثيرة رئيسية في سبيل عرض ناحية أو وجهة نظر معينة في علم النفس عرضاً أكثر شمولاً وفعالية هو أيضاً غير كاف، لأن في هذا الحل لا يُعرض كثير من الميادين الرئيسية على الطالب إطلاقاً.

إن مكتبة أصول علم النفس الحديث كانت المحاولة الأولى في إصدار مجموعة من الكتب الصغيرة الحجم التي تتناول موضوعات أساسية مختلفة، يكتب كلاً منها عالم متخصص كفء. ثم أخذ هذا الإتجاه في التأليف في علم النفس يتزايد بعد ذلك. ولقد كنا متأثرين في إصدار هذه السلسلة من الكتب بفكرة تزويد القائمين بتدريس المقررات العامة في علم النفس بمادة تكون أكثر مرونة من المادة الموجودة في الكتب الحالية الكبيرة ذات الطابع الموسوعي،

وعرض موضوع واحد في كل كتاب عرضاً أكثر عمقاً مما لا يتوفر في كتب المداخل التي لا تفرد لهذه الموضوعات عادةً حيزاً كبيراً.

إن أول كتاب في هذه المكتبة ظهر عام ١٩٦٣، وآخرها ظهر في عام ١٩٦٧. ولقد بيع من هذه الكتب أكثر من ربع مليون نسخة مما يشهد على استخدام هذه الكتب استخداماً واسع النطاق في تدريس علم النفس. وقد استخدم بعض كتب هذه المكتبة ككتب إضافية، واستخدم بعضها كالكتاب المقرر في كثير من مقررات المرحلة الأولى الجامعية في علم النفس، والتربية، والصحة العامة، وعلم الاجتماع. كما استخدمت مجموعة من كتب هذه المكتبة ككتب مقررة في المقررات التمهيدية في علم النفس العام في المرحلة الأولى الجامعية. وقد ترجم كثير من هذه الكتب إلى ثماني لغات هي الهولندية، والعبرية، والإيطالية، واليابانية، والبولندية، والبرتغالية، والأسبانية، والسويدية.

ولوجود اختلاف كبير في زمن نشر هذه الكتب، ونوع محتوياتها فإن بعضها يحتاج إلى مراجعة، بينما بعضها الآخر لا يحتاج إلى ذلك. ولقد تركنا اتخاذ هذا القرار إلى مؤلف كل كتاب، فهو الذي يعرف جيداً كتابه من حيث علاقته بالوضع الحالي للميدان الذي يتناوله الكتاب. وسيظل بعض هذه الكتب بدون تغيير، وبعضها سيعدل تعديلاً طفيفاً، وبعضها سيعاد كتابته كلية. ولقد رأينا أيضاً في الطبعة الجديدة لهذه المكتبة أن يحدث بعض التغيير في حجم بعض الكتب وفي أسلوبها لتعكس بذلك الطرق المختلفة التي استخدمت فيها هذه الكتب كمراجع.

لم يكن هناك من قبل على الإطلاق اهتمام شديد بالتدريس الجيد في كلياتنا وجامعاتنا كما هو موجود الآن. ولذلك فإن توفير الكتب القيمة والمكتوبة جيداً والمثيرة للتفكير والتي تلقي ضوءاً على البحث المتواصل المثير عن المعرفة يصبح متطلباً أساسياً. ويصبح ذلك ضرورياً على وجه خاص في مقررات المرحلة الأولى الجامعية حيث يجب أن تكون في متناول يد عدد كبير من الطلبة كتب تمدهم بقراءات مناسبة. إن مكتبة أصول علم النفس الحديث تمثل محاولتنا المستمرة لتزويد مدرسي الكليات بالكتب المقررة التي يمكننا تأليفها.

ريتشارد س. لازاروس
(المشرف على إصدار المكتبة باللغة الانجليزية)

المحتويات

تصدير الطبعة العربية

إن تعقيدات الحياة اليومية في العصر الحديث، وكثرة التخصصات المهنية المختلفة، وتعدد فروع الدراسات العلمية والأدبية والفنية، كل ذلك جعل من الصعب على الفرد أن يقوم باتخاذ قرار واضح وسليم حول نوع التعليم المناسب له، أو حول نوع المهنة المناسبة له؛ كما جعل من الصعب على المؤسسات التربوية والمهنية اتخاذ قرارات واضحة وسليمة حول الأفراد لاختيار المسار المناسب لهم من حيث نوع التعليم أو المهنة كما أنه من الصعب على المختصين في العلاج النفسي، في بعض الحالات، اتخاذ قرارات واضحة حول تشخيص نوع المرض النفسي أو العقلي، وتحديد نوع العلاج النفسي المناسب. ومن هنا نشأت الحاجة إلى ابتكار الاختبارات والمقاييس النفسية واستخدامها في قياس القدرات العقلية وسمات الشخصية من أجل مساعدة الأفراد على اتخاذ قرارات حول أنفسهم والتخطيط لمستقبلهم، ومساعدة المؤسسات التربوية والمهنية ومؤسسات العلاج النفسي على اتخاذ قرارات سليمة حول الأفراد الذين تتولى أمر تعليمهم أو توجيههم أو علاجهم.

ومع أن الاختبارات والمقاييس النفسية، وبخاصة اختبارات الذكاء، أصبحت الآن شائعة الاستعمال في كل هذه الميادين التي أشرنا إليها سابقاً، إلا

أن كثيراً من الناس لا زالت تنقصهم المعرفة الصحيحة والدقيقة بطبيعة هذه الاختبارات، وخصائصها، وحدودها، وكيفية تقييم نتائجها تقييمًا صحيحاً.

وهذا الكتاب الذي نقدمه اليوم إنما هو محاولة جادة وموفقة لعرض موضوع «الاختبارات والمقاييس النفسية» عرضاً واضحاً دقيقاً، يوضح طبيعتها، ويشرح مفاهيمها الأساسية، ويبين خصائصها وشروطها، ويشرح تطبيقاتها العملية المختلفة. ولا شك، أن هذا الكتاب سيكون مفيداً سواء للطلبة أو للعاملين في جميع الميادين التي تتصدى للحكم على الأفراد أو لاتخاذ قرارات في شأنهم. كما أنه سيكون مفيداً أيضاً للقارىء العادي إذ سيمدّه بفكرة واضحة عن الاختبارات والمقاييس النفسية مما يجعله أكثر فهمًا لطرق استخدامها، وأكثر دقة في تقييم نتائجها.

وأخيراً، يسرني أن أتقدم للزميل الأستاذ الدكتور سعد عبد الرحمن بجزيل الشكر والثناء على ما بذل من جهد موفق في ترجمته لهذا الكتاب الذي سيلقى، بلا شك، ترحيباً من جميع المشتغلين بالقياس النفسي في ميادين علم النفس والتربية والتوجيه المهني والاختيار المهني والعلاج النفسي.

محمد عثمان نجاتي

١٩٨٢/٨/٢

تصــديرُ المؤلّفــة

إن مراجعة كتاب مدرسي بعد فترة سبع سنوات من صدوره تتيح فرصة ممتازة لتقييم التقدم والتغير اللذين حدثا في الميدان الذي يتناوله الكتاب. وإذا ما نظرنا إليه، آخذين في اعتبارنا التغيرات الهائلة التي حدثت في جوانب أخرى من الميدان التربوي خلال الستينات، لوجدنا أن السمات الرئيسية لهذا الميـدان الخاص الذي يتناوله هذا الكتاب كان نوعاً من الثبات النسبي. لقد كانت هناك تغيرات جديدة كثيرة، وأساليب إحصائية جديدة، واختبارات جديدة، وتقنينات جديدة لاختبارات قديمة، ولكن كانت هناك تغيرات قليلة في المفاهيم والمبادىء.

إن الازدياد الهائل في حجم الاستخدامات العامة للآلات الحاسبة وفي الحنكة في استخدامها يعمل على تغيير كل تكنولوجيا البحث الكمي والقياس النفسي. إن عمليـات التحليـل العـامـلي، وتحليـل البنـود، ومعـادلات الانحراف وما شابه ذلك يمكن أن تجري بصورة أسهل وأسرع مما كانت تجري به عمليات الارتباطات البسيطة من قبل. وهناك أحد التطورات الجديدة في المفاهيم يبدو وشيكاً في الأفق قد يؤدي يوماً ما إلى تغير كبير في طريقة تصميمنا

للاختبارات. فالآلات الحاسبة الآن تجعل من الممكن جعل الاختبارات فردية - أي نختار لكل فرد مجموعة مُثلى من البنود من أجل زيادة الدقة في تقييم إمكاناته وتحصيله. ولكن لم يحن بعد الوقت الذي يصبح فيه مثل هذا الاتجاه في الاختبارات هو الاستخدام السائد.

إن أوضح فرق بين عام ١٩٦٣[1] وعام ١٩٧٠[2] هو في اتجاه الرأي العام نحو القياس النفسي. ويبدو الأمر في الوقت الحالي بصورة غير مرضية بدرجة كبيرة بالمقارنة بما كان عليه الأمر في عام ١٩٦٣. فقد أثيرت الشكوك حول الفائدة الاجتماعية لمشروع الاختبارات بأكمله. فالبعض يهتم بإمكان حدوث تحيز في اختبارات الاستعدادات بحيث يؤدي ذلك إلى إبعاد الفقير وغير المتعلم من فرص تحسين حالاتهم. والبعض الآخر يخشى من أن أنواع الأسئلة التي كانت تسأل في استبيانات الشخصية تنتهك حق الفرد في السرية وهناك أيضاً بعض الناس الآخرين الذين يقلقون من عواقب بنوك المعلومات التي تصبح فيها درجات الفرد في الاختبارات جزءاً دائمًا من سجله، من الممكن استرجاعه في سنوات لاحقة، وربما يكون ذلك في غير صالحه.

إن هذه مشكلات حقيقية. وقد حاولت أن أتناولها في الأجزاء المناسبة في هذه الطبعة المراجعة من الكتاب، للتأكد من أن المستهلك العام للمعلومات التي تمدنا بها الاختبارات، وهو الذي له يكتب هذا الكتاب، لن يغفل عنها.

وفي مواجهة الاضطراب الموجود في الوقت الحالي، فإن كثيراً من المفكرين المتطرفين بدرجة شديدة أو غير شديدة ينادون بفكرة التخلص من الاختبارات النفسية كلية. ويبدو لي أن الأمر الأكثر معقولية هو أن نحاول تحسين الاختبارات، وأن نحاول استخدامها بفهم. وإني لأرجو أن يستطيع هذا الكتاب أن يساعد قرّاءه على فعل ذلك.

ليونا ا. تايلر

---

(١) وقت ظهور الطبعة الأولى من الكتاب (المترجم).

(٢) وقت ظهور الطبعة الثانية من الكتاب (المترجم).

الفصل الأول

طبيعة ووظيفة القياس في علم النفس

إن الإنسان في حاجة لأن يفهم طبيعته، فقد أصبح من نافلة القول أن المشكلات الرئيسية التي يواجهها عالمنا اليوم إنما هي مشكلات بشرية، وأن المعوقات الأساسية التي تمنع التقدم إنما هي معوقات بشرية. وفضلاً عن ذلك، فإن المجتمع إذْ قد تطور الآن إلى نظام معقد من العلاقات والأدوار المتخصصة، فقد أصبح الناس في حاجة إلى أدوات جديدة لتساعد الأفراد لكي يجدوا أماكن ملائمة لهم في هذا البناء الكلي.

وهذه الحاجات الملحة في هذا العصر هي التي جعلت بالضرورة أن يصبح علم النفس علمًا. فالأوصاف الأدبية والفلسفية لطبيعة الإنسان التي كانت تكوّن نواة علم النفس في الحقب الماضية قد حلت محلها في هذا العصر التعليلات والتفسيرات التي تقوم على البحث التجرُبي. ولكي يقوم علماء النفس بما يريدون من بحوث فإن عليهم أن يعالجوا متغيرات الدراسة بصورة كمية، وهكذا أصبح القياس علامة مميزة لعلم النفس الحديث.

وهذه الحقيقة لا ترضي الكثيرين من طلاب علم النفس، لأنهم أقبلوا

على دراسة هذا العلم ليزيدوا فهمهم للأسباب التي تفسر لماذا هم والناس الذين يعرفونهم يشعرون ويتصرفون على هذا النحو، ولكنهم يجدون أنفسهم في مواجهة بعض المفاهيم التي تبدو بعيدة كل البعد عن الاهتمامات الإنسانية: مثل الانحرافات المعيارية، والتوزيعات التكرارية، والاحتمالات. وليس من المستغرب أن يعترض الطلاب ويشكون من المتطلبات التي توضع لمقررات علم النفس التي يختارونها للدراسة. فهم يتساءلون عن علاقة وأهمية المفاهيم الرياضية والقوانين الكمّية لشخص يهدف إلى الحصول على المعرفة المفيدة في علم النفس.

والحقيقة هي أن التفكير الكمي مَعْلَم أساسي أكثر منه سطحي في علم النفس في الوقت الحاضر. فالتقدم الذي تم خلال القرن الماضي كان لا يمكن أن يتم بدون هذا النوع من التفكير. ففي المقام الأول، نجد أن الطرق الكمية تساعدنا على الوصول إلى استنتاجات دقيقة من التجارب. فعند القيام بتجربة ما، فإننا نطبق منهجاً خاصاً على الحيوان أو الإنسان ونلاحظ التأثيرات الناتجة. ولكنه من الواضح لأي ملاحظ دقيق أن ردود الأفعال للمواقف المثيرة تختلف كثيراً من فرد إلى فرد. فعلى سبيل المثال، عندما يقصف الرعد بصورة مفاجئة فقد نجد البعض يجري إلى النافذة ليشاهد ومضات البرق، والبعض الآخر يخفي وجهه بين يديه. وكذلك في التجارب، نجد أن ردود الأفعال تتباين كثيراً. وإذا لم نجد إجابة ما على الأسئلة التي يكون أولها (كَمْ)، فإن نتائج كل التجارب التي تجري على أكثر من فرد واحد تصبح متناقضة ومتضاربة. وإذا لم نقم بدراسة أكثر من فرد واحد، فإننا لا نتوقع أن نعرف كثيراً عن الطبيعة البشرية.

فعلى سبيل المثال، عند دراسة عملية التعلم، فقد يرغب الإخصائي النفسي في أن يعرف هل من الأفضل أن يقدم للطلاب موضوعات الدراسة بمنطق بسيط بحيث لا يقع الطلاب في الأخطاء، أو أن يقدم هذه الموضوعات بمنطق آخر يسمح بالوقوع في الخطأ، ثم يقوم بتصحيح هذه الأخطاء. (مثل هذه الموضوعات يصبح على جانب كبير من الأهمية عندما نتولى المهمة العملية الخاصة ببناء برامج تعليم يتم بمساعدة الحاسب الآلي). فعند تصميم تجربة للإجابة على هذا التساؤل فإن الإخصائي النفسي قد يستخدم طرقاً عديدة لقياس أداء المفحوصين في نواحٍ متعددة على قدر الإمكان. فإذا كان هذا

الإخصائي يتوقع أن المفحوصين لن يستجيبوا جميعاً بطريقة واحدة لظروف التجربة، فإنه لن يدهش إذا وجد أن بعض المفحوصين يسجل درجات أعلى تحت الظروف التجريبية (أ)، وآخرين يسجلون درجات أعلى تحت الظروف (ب)، وآخرين يسجلون درجات عالية بدرجة متساوية تحت كل من الظروف (أ) و (ب). وقد تكون النتيجة أن الظروف (أ) تساعد بعض المفحوصين على التعلم بسرعة جداً، ولكن على حساب الاحتفاظ بما يتعلمونه، بينما تكون الظروف (ب) تستغرق وقتاً أطول في التعلم، ولكنها تسهّل عملية الاحتفاظ. ومن الواضح أن الإخصائي النفسي قد يفكر في الكثير من الاحتمالات المعقدة عندما يصمم التجربة.

ولهذا السبب فإن الإخصائي النفسي قد يلجأ إلى أنواع مختلفة من المقاييس ليتمكن من تقييم هذه الاحتمالات. ولذلك، فربما يقيس أولاً الزمن الذي يستغرقه كل مفحوص ليصل إلى مستوى معين من الأداء. وربما يطبق اختباراً في التفكير المنطقي مباشرة بعد إجراء التجربة، ثم مرة أخرى بعد مرور شهر عليها. وربما يصمم اختبارات لأنواع أخرى من القدرات العقلية التي يظن أن لها علاقة بالتفكير المنطقي، مثل حل المشكلات بطريقة عملية، أو اكتشاف الأخطاء في الدعاية. وفي كل مجموعة من الدرجات يقوم بمقارنة متوسطي المجموعة (أ) والمجموعة (ب)، مستخدماً في ذلك الطرق الإحصائية التي تساعده على الوصول إلى استنتاج يمكن صياغته بمقتضى فكرة الاحتمالات. إن ما يقرره، إذن، قد يأخذ هذه الصيغة: «في هذه العينة وجد أن الطريقة (ب) أفضل من الطريقة (أ)، وأن هناك احتمالاً أقل من خمس مرات في كل مئة مرة لأن يظهر مثل هذا الفرق بين الطريقة (أ) والطريقة (ب) في عينة مأخوذة من مجتمع لا توجد بين أفراده فروق حقيقية». وبدون هذه المعالجة الكمية للنتائج كان لا يمكن للباحث أن يصل إلى أية استنتاجات على الإطلاق؛ ولكن بهذا النوع من المعالجة الكمية أمكنه أن يصل إلى قرار حول احتمالٍ يمكن أن يستند عليه في قرارات تالية: مثل المضي في تطبيق إحدى الطريقتين، أو القيام بتجربة أخرى.

وهكذا فإن المقاييس تساعد الإخصائيين النفسيين في اتخاذ قرارات حول معنى ما يقومون به من بحوث. وتيسّر المقاييس أيضاً إصدار قرارات خاصة بالأفراد. وقد بدأ الإخصائيون النفسيون في بناء الاختبارات العقلية لأنه كانت

هناك حاجة لمثل هذه الأدوات التي تساعد في اتخاذ هذا النوع من القرارات. ويعتبر مقياس ألفرد بينيه Alfred Binet في عام ١٩٠٥ الأصل الأول لكل اختبارات الذكاء في الوقت الحاضر. وقد أعد بينيه هذا الاختبار كوسيلة لمساعدة المسؤولين عن التعليم في باريس للتعرف على الأطفال الذين لا يستطيعون الاستفادة من برامج المدارس العادية مهما بذلوا من جهد. ومهما يقال عن سوء استخدام اختبارات الذكاء منذ عهد بينيه حتى وقتنا الحاضر - وهناك العديد من الانتقادات الجادة والصحيحة التي يمكن توجيهها إليها - إلا أن هذه الاختبارات قد ساعدت المدرسين في حالات كثيرة على التمييز بين الطفل الغبيّ والطفل الكسول بطريقة أفضل مما لو اتبعوا أسلوباً آخر.

ولنأخذ مثالاً آخر. ففي خلال الثلاثينات من هذا القرن واجهت مكاتب التوظيف العام في جميع أنحاء الولايات المتحدة الأمريكية مهمة تبدو كما لو كانت مستحيلة التنفيذ. فآلاف الأفراد فقدوا وظائفهم بسبب الكساد الاقتصادي، وآلاف من الأولاد والبنات يتخرجون كل عام من المدارس باحثين عن عمل. وكانت المعضلة الرئيسية، بالطبع، هي عدم وجود فرص كافية للعمل. ولكن كانت هناك مشكلة أخرى لا تقل خطورة، وهي أن الكثير من الباحثين عن العمل عندما يلتحقون بأعمال ما، إن وجدت هذه الأعمال، يجدون أنهم قد ألحقوا بأعمال لا تتوفر لديهم خبرة سابقة فيها. وهذا ما دفع إلى البدء بتنفيذ برنامج واسع لبناء اختبارات في الاستعدادات المهنية صممت خصيصاً للتنبؤ بإمكانية نجاح الفرد في أية مهنة قبل الالتحاق بها. وكانت نتيجة هذا البرنامج بطارية اختبارات الاستعدادات العامة[1] - التي يستغرق تطبيقها ساعتين ونصف - التي جعلت من الممكن تقييم مدى صلاحية الأفراد لمئات من الأعمال المختلفة.

وهناك نوع آخر من القرارات التي ساهم فيها الإخصائيون النفسيون والتي مهدت لتكوين مجموعة جديدة من الاختبارات وأساليب القياس. فقد تزايد اهتمام كثير من الناس بالصحة العقلية، كما تزايد الوعي بإمكانية معالجة الأمراض النفسية، وبالتالي ظهرت الحاجة إلى اختبارات الشخصية. فعندما يأتي مريض إلى مصحة عقلية أو عيادة فإن من المفيد أن يعرف منذ البداية أي نوع

(١) General Aptitude Test Battery (G A T B).

من الأشخاص هو وكيفية استجابته للعلاج. خذ على سبيل المثال حالة الفرد ل. ج. فهل تشخص حالته على أنها «فصام» أو «استجابة القلق»؟ ولكن الطبيب النفسي يبحث عن المعلومات التي تساعده على معرفة هل يستجيب ل. ج. للعلاج أفضل إذا أدخل المستشفى، أم تكون استجابته أفضل للعلاج في العيادة النفسية القريبة من المنطقة التي يسكن فيها. وإذا أدخل المستشفى، فهل يلحق بإحدى الجماعات العلاجية، أو هل من الأفضل أن يتم علاجه بصورة فردية؟ مثل هذه القرارات المعقدة يصعب اتخاذها بدرجة كبيرة من اليقين، ولكن اختبارات الشخصية المتنوعة تساعد على اتخاذها.

وأساليب القياس لها أهميتها أيضاً في المواقف التي يجب فيها على الأفراد أن يتخذوا قرارات تخص حياتهم. فقرارات الفرد عن نفسه هي، من بعض النواحي، أكثر تعقيداً من قراراته عن الآخرين. فكثير من العوامل الداخلة في هذه القرارات لا يمكن عادةً معالجتها معالجة كمية - فالعوامل التي تكون شخصية جداً مثل أن يعرف السيد (هـ) أن أمّه سوف يخيب أملها إلى حد كبير إذا لم يدخل في سلك رجال الدين، أو شعور السيد (م) تجاه عمل ما يتطلب منه أن يتغيب عن بيته لفترات طويلة من الزمن. ولكن وجود إجابات كمية لبعض الأسئلة كثيراً ما يسهل هذه المواقف إلى الدرجة التي تمكن من اتخاذ القرارات المُرْضية. فالدرجة على اختبار ما، على سبيل المثال، سوف تساعد الطالب في السنة النهائية من دراسته الذي يحاول اتخاذ قرار حول متابعة دراسته العليا أو الاتجاه إلى العمل، وذلك عن طريق مقارنته - من حيث القدرات العقلية المركبة - بطلبة الدراسات العليا المتفوقين في ميدان تخصصه. والدرجة على اختبار آخر سوف تساعده على معرفة مدى التشابه بين اهتماماته واهتمامات رجال الأعمال أو الدارسين الأكاديميين في ميدان تخصصه. وهكذا فإن القياس النفسي يسهم في اتخاذ مثل هذه القرارات يومياً في مكاتب الاستشارة في المدارس العليا والكليات الجامعية.

وعلى ذلك، فإن وسائل القياس والطرق الإحصائية المصاحبة لها قد دخلت في النسيج الكلي لعلم النفس العام وعلم النفس التطبيقي. وسوف نناقش الآن في شيء من التفصيل الأفكار والأدوات التي يتم استخدامها. ويجب أن نقول إن المهارة الحقيقية في استخدام هذه الطرق والوسائل تتطلب تدريباً طويلاً، ولكن الفهم العام للمبادىء الأساسية هو في متناول كل شخص - مثل

الآباء، والمعلمين، ورجال الأعمال، أو حتى المواطن العادي.

مستويات القياس

إن كلمة القياس، كما يستخدمها علماء النفس، تشمل مجموعة كبيرة من العمليات. إن الشيء الوحيد المشترك بين جميع هذه العمليات هو استخدام الأرقام. فالتعريف العام جداً الذي يمكن أن نضعه للقياس هو ببساطة أن القياس يعني تحديد أرقام حسب قواعد معينة.

وهذه القواعد ليست ذات طبيعة ضيقة ومحددة كما كثيراً ما يظن البعض من ذوي المعلومات الرياضية المحدودة. فإذا لم نفكر جيداً في ماهية القياس فإنه من المحتمل جداً أن نفترض أن العمليات الحسابية الأولية ـ وهي الجمع والطرح والضرب والقسمة ـ يجب أن تطبق على كل أنظمة القياس. وعلى ذلك، فقد يكون رد الفعل الأولي هو أن نستنتج أن القياس مستحيل إذا لم يكن من الممكن تطبيق بعض العمليات الحسابية. فلنأخذ كـــل ميدان قياس الذكاء على سبيل المثال. فمنذ بداية استخدام هذه الاختبارات، نجد أن ذوي الفكر الناضج كثيراً ما يقولون إنه لا معنى للقول بأن الطفل إذا كانت نسبة ذكائه ١٥٠ فإنه يكون أذكى مرتين من طفل نسبة ذكائه ٧٥. إن اختلاف الأطفال بعضهم عن بعض في الاستجابة لمواقف التفكير المجرد حقيقة واقعة، ولكن ليس هناك أي مبرر في سلوكهم يدعو إلى القول بأن هذا الاختلاف يمكن التعبير عنه كحاصل قسمة = ٢، أو كرقم كسري مثل $\frac{١}{٢}$ ، أو كنسبة مثل ٢ : ١. فالحقيقة أنه لا معنى لقسمة نسبة ذكاء ما على نسبة ذكاء أخرى..

وبمجرد أن انتبه الإخصائيون النفسيون إلى هذه الخاصة في بعض الأرقام التي كانوا يستخدمونها، أدركوا أن هناك مواقف مناظرة لذلـك خارج علم النفس. خذ درجة الحرارة، مثلاً، كما نقيسها عادة سواء بالتدريج الفهرنهايتي أو المئوي. فإذا انخفضت درجة الحرارة من ٨٠° (ف) أثناء النهار إلى ٤٠° (ف) أثناء الليل، فإننا لا نقول أن دفء منتصف الليل يساوي نصف دفء الظهيرة، وذلك لأن الصفر في نظام الفهرنهايت هو رقم تحكّمي، وهو لا يعني في الحقيقة عدم وجود الدفء تماماً. ولذلك، فإن أرقام الدرجات التي تعلو الصفر لا يمكن أن تعالج بنفس الطريقة التي تعالج بها الأرقام في مقياس للطول أو الوزن. ومن

الواضح أن مقاييس الذكاء أشبه بمقاييس الحرارة منها بمقاييس الطول.

وقد تمكن ستيفنز[1] S. S. Stevens من تقديم تصور عام للأنواع أو المستويات المختلفة للقياس التي كانت ذات فائدة كبرى لعلماء النفس. وبناء على هذا التصور أو التنظيم فإنه يمكن أن نقسم طرق استخدام الأرقام إلى أربعة أنواع. كل نوع من هذه الأنواع المختلفة من القياس له قواعده وحدوده الخاصة. كما أن لكل منها طرقاً إحصائية معينة مناسبة له.

وتنظيم ستيفنز، كما هو مبين في الشكل رقم ١، يرتب مستويات القياس هذه بناء على مدى تطبيق العمليات الحسابية المألوفة. فالمستوى الأول أو الأدنى في هذه السلسلة هو مقياس **التصنيف**. وفي هذا النوع من المقاييس نستخدم الأرقام للدلالة فقط على **الفئات** التي ينتمي إليها الأفراد أو الأشياء. وهذه يمكن أن تكون تصنيفاً لأفراد نوع واحد، مثل ترقيم ملابس لاعبي كرة القدم لتمييزهم؛ أو قد تكون تصنيفاً **لمجموعات** من الأفراد، مثل إعطاء مجموعة الرجال الرقم ١، ومجموعة النساء الرقم ٢. والعملية الحسابية الوحيدة التي يمكن أن تطبق في مقاييس التصنيف هي عملية العدّ، أي مجرد تعداد الأفراد في كل صنف أو مجموعة. وهذه الأعداد المميزة لا يمكن أن تجري عليها أي من العمليات الحسابية الأربع: الجمع أو الطرح أو الضرب أو القسمة.

والمستوى الثاني من مستويات القياس هو مقياس **الترتيب**. ونحن نستخدم هذا النوع من المقاييس عندما نريد أن نرتب الأفراد في سلسلة تبدأ من الأقل وتنتهي بالأعلى بناء على الخاصة المطلوب قياسها، ولكن لا يمكننا أن نقول بدقة كم يكون الفرق بين أي اثنين من هؤلاء الأفراد. فعندما تقوم لجنة ما بترتيب خمسة من المتقدمين للمنح الدراسية بناء على تقديرهم العام فإنها تستخدم في ذلك مقياس الترتيب. وكثيراً ما تستخدم الدرجات المئينية في التعبير عن نتائج اختبارات الطلاب، وحيث أن هذه الدرجات هي نوع من الترتيب، فإنها أيضاً تمثل مقياس ترتيب. والعمليات الحسابية المعروفة ـ الجمع والطرح والضرب والقسمة ـ لا يمكن أن تستخدم على وجه صحيح في حالة مقياس الترتيب، ولكن الطرق الإحصائية التي تقوم على فكرة الترتيب تكون صالحة للاستخدام.

---

(١) انظر فصل: «الرياضيات، والقياس، والسيكوفيزياء» في كتابه:
Handbook of Experimental Psychology. New york: John Wiley and Sons, 2nc., 1951.

فمن الممكن، على سبيل المثال، أن نعيّن ما إذا كانت هناك علاقة بين رتب مجموعة من الأطفال على مقياس للشعبية ورتبهم على مقياس للاعتماد على الغير.

والمستوى الثالث من المقاييس هو مقياس الوحدات المتساوية. إن ما يميز مقياس الوحدات المتساوية من مقياس الترتيب هو أنه يمكننا من معرفة مقدار المسافة بينن شيئين أو شخصين. ومعظم الاختبارات المدرسية هي من هذا النوع. وعلى الرغم من أن هناك بعض التساؤلات حول الخصائص القياسية لهذه الاختبارات، إلا أن المعلمين عادة ما ينظرون إلى الدرجات على اختبارات

المثال	حدود المستوى	المستوى
	لا توجد حدود ـ فكل العمليات الحسابية والأساليب الإحصائية يمكن استخدامها.	٤- مقاييس النسبة كل رقم يمكن أن ينظر إليه باعتباره قياساً لمسافة ما من نقطة الصفر.
	بالإضافة إلى الطرق المذكورة أدناه فإنه يمكن استخدام الجمع والطرح والأساليب الإحصائية القائمة على هذه العمليات الحسابية. أما الضرب والقسمة فلا يمكن استخدامهما	٣- مقاييس الوحدات المتساوية الفترات أو المسافات بين كل رقم والرقم الذي يليه متساوية ولكن ليس من المعروف كم يبعد أي رقم منها من الصفر
	بالإضافة إلى الطرق المذكورة أدناه فإنه يمكن استخدام أساليب الترتيب والأساليب الإحصائية الأخرى القائمة على التقسيمات التي تشير إلى «أكبر من» أو «أقل من»	٢- مقاييس الترتيب تشير الأرقام إلى رتبة أو ترتيب.
	إن العمليات الحسابية المسموح باستخدامها فقط هي العد والأساليب الإحصائية القائمة على العد.	١- مقاييس التصنيف تستخدم الأرقام لتسمية الأشياء وإثباتها وتصنيفها.

شكل (١): مستويات القياس

المقررات الدراسية باعتبارها مقاييس وحدات متساوية. فالمعلمون يعتقدون أنه من المعقول مقارنة درجات طالبين، ثم الحكم بأن درجات الطالب (ج) تقل بمقدار ١٦ وحدة عن درجات الطالب (ب). ولكن مقاييس الوحدات المتساوية لها عيب واحد هام، وهو أنه ليس لها نقطة صفر حقيقي. حقيقة أنه قد يحدث أحياناً أن يحصل الطالب على درجة «صفر» في امتحان للهندسة أو الأدب الإنجليزي. ولكن هذا لا يعني أن هذا الطالب ليست لديه معلومات البتة في أي من هذين الموضوعين. وعلى الرغم من أن هدف الاختبار هو قياس المعرفة بمحتوى المادة، إلا أنه ليس من الضروري في الواقع أن نحدد معنى «الصفر في المعرفة» حتى يتم قياس المعرفة بهذا المحتوى، إذ أننا سنستخدم الاختبار أساساً لمقارنة الأفراد بعضهم ببعض.

ورغم أنه يمكننا أن نجمع ونطرح الدرجات على مقياس الوحدات المتساوية، إلا أن هناك عملية حسابية لا يمكن أن تجوز ـ وهي عملية قسمة درجة على درجة أخرى، إذ أن عملية القسمة تتطلب وجود درجة صفر حقيقية. ولكن لماذا كان ذلك لا يجوز؟ لنفرض أن الطالب (ل) حصل على ٨٠ درجة في اختبار ما، وأن الطالب (م) حصل على ٤٠ درجة في نفس الاختبار. ولنفرض أن الأستاذ عند إعادة صياغة هذا الاختبار أضاف عشرة أسئلة سهلة بحيث يمكن أن يجيب عليها كل من الطالبين (ل) و (م) إجابة صحيحة. في هذه الحالة تصبح درجة الطالب (ل) ٩٠، بينما تصبح درجة الطالب (م) ٥٠. وسيكون الفرق بين درجتي (ل) و (م) هو ٤٠ درجة في الحالتين، ولكن النسبة بين درجتيهما سوف لا تكون واحدة، حيث تكون في المرة الأولى $\frac{٨٠}{٤٠}$ = ٢، وفي الثانية $\frac{٩٠}{٥٠}$ = ١,٨. ففي أي اختيار، إذن، لا يمكن أن نعرف ما إذا كانت معلومات فرد ما هي ضعف معلومات الفرد الآخر، أو ثلاثة أمثالها، أو تعادلها مرة ونصف. ولكن إذا سلمنا بأن كل سؤال من أسئلة الاختبار يعتبر مؤشراً جيداً للمعرفة في هذا الميدان أو ذاك، فإننا لا ننتهك قواعد الرياضيات والمنطق عندما نطرح درجة من درجة أخرى، أو عندما نجمع مجموعة من الدرجات ثم نستخرج المتوسط. فالوسائل الإحصائية التي تستخدم لتحويل الدرجات الخام في أي اختبار إلى أنواع مختلفة من الدرجات المشتقة ـ وهذه سوف نناقشها فيما بعد ـ تعتمد أساساً على عمليتي الجمع والطرح.

والمستوى الرابع من مستويات القياس ـ وهو الأعلى ـ هو مقياس النسبة، وفيه يمكن استخدام جميع العمليات الحسابية: الجمع والطرح والضرب والقسمة. ومقاييس النسبة لها جميع خصائص مقاييس الوحدات المتساوية، بالإضافة إلى خاصة مميزة وهي وجود صفر حقيقي. وربما كان هذا النوع من المقاييس مألوفاً لدينا أكثر من الأنواع الأخرى لأن كل الأبعاد الطبيعية المعروفة ـ الطول والوزن والحجم ـ إنما تقاس بهذا المقياس. فعلى ميزان القصاب الأمين يعني الصفر أنه لا يوجد أي لحم على الإطلاق في كفة الميزان. وعندما نشتري منه نكون على ثقة من أن قطعة اللحم زنة ستة أرطال أثقل بمقدار أربعة أرطال من قطعة أخرى وزنها رطلان فقط. ويمكن أن نقول أيضاً أن القطعة الأولى تعادل وزن القطعة الثانية ثلاث مرات. ومقياس النسبة حسب التسمية يدل على أنه يمكننا أن نقسم عدداً على عدد آخر، أو أن نعبّر عنهما في صيغة نسبة.

وكثير من المقاييس التي نستخدمها في علم النفس لا يمكن اعتبارها مقاييس نسبة، وإن كان عدد قليل منها يمكن اعتباره كذلك. وهذه المقاييس القليلة توجد عادة عندما يكون في الإمكان قياس خاصة عقلية عن طريق وحدات فيزيائية من نوع ما. فعندما نقيس زمن الرجع، على سبيل المثال، فنحن نستخدم وحدات الزمن المعروفة وهي الثواني وكسور الثواني. فإذا أردنا أن نعين سرعة استجابة سائق السيارة للضغط على الكوابح (الفرامل) عندما يظهر الضوء الأحمر، فإنه يمكن معرفة ذلك باستخدام مقياس النسبة. وعلى ذلك، فإذا كان السائق (ج) يحتاج إلى $\frac{٥}{١٠}$ من الثانية، في حين أن السائق (ب) يحتاج إلى $\frac{٣}{١٠}$ من الثانية، فإنه يمكننا أن نستخدم النسبة بين سرعة كل منهما وهي ٣ : ٥ لكي نصف كم تزيد سرعة (ب) عن سرعة (ج). وحتى عندما نستخدم مقياس النسبة، كما في هذه الحالة، فقد لا نرغب في استخراج نسب أو قسمة عدد على عدد آخر. وعلى العموم فإن كل الطرق المستخدمة في المستويات الأولى من القياس يمكن أن تستخدم في المستويات الأعلى. وفي التطبيق العملي لدراسات زمن الرجع، على سبيل المثال، فإن ما يريد الباحث أن يعرفه عن السائقين (ب) و (ج) قد يكون هو مدى مقارنة كل منهما بالمعايير والمستويات التي تميز السائق الماهر. ومثل هذه المقارنات تتطلب عملية الطرح فقط، ولا تستدعي استخراج النسبة.

ومنذ تقدم ستيفنز بهذا النظام لتصنيف مستويات القياس، لا زال الحوار دائراً على نطاق واسع حول ما هي الطرق الصحيحة والطرق غير الصحيحة لمعالجة أنواع معينة من البيانات في البحوث النفسية. ولسوء الحظ ليست كل مجموعات الأرقام يمكن أن توضع بوضوح تحت صنف أو آخر من هذه الأصناف الأربعة كما هو الحال بالنسبة للأمثلة المستخدمة في الشكل رقم ١. فأرقام التليفونات، مثلاً، من المحتمل أن تمثل مقياساً للتصنيف لغرض التعريف فقط. في حين أن أرقام المناطق قد تمثل مقياساً للترتيب للدلالة على بعد أو قرب هذه المنطقة أو تلك من جهة ما. فأي نوع من العمليات العددية يمكن تطبيقه على أرقام المناطق؟.

وإن تقرير ما إذا كانت مجموعة من القياسات تقع تحت فئة مقاييس الترتيب أو مقاييس الوحدات المتساوية لهو أيضاً أمر صعب في بعض الأحيان. فالدكتور (س)، على سبيل المثال، يصل إلى استنتاج مؤداه أن القلق والإبداع مرتبطان معاً بدرجة ما، وهو يؤيد هذا الاستنتاج بأن معامل الارتباط بين مقياسين لهاتين الخاصتين يصل إلى ٠,٣٦ في مجموعة درسها من طلاب المدرسة الثانوية. والدكتور (ص) يعترض على استنتاج الدكتور (س) على أساس أنه ليس هناك ما يؤكد أن القلق والإبداع قد تم قياسهما بوساطة مقياسين للوحدات المتساوية، وبالتالي ليس من الصحيح حساب معامل الارتباط بين الدرجات بالطريقة التي اتبعها الدكتور (س).

وهناك طريقة تتبع لتجنب الخلافات التي من هذا النوع وهي أن نعتبر النظم الرياضية التي تطبق على البيانات بمثابة نماذج تشبه النماذج التي يعدها الجيولوجيون لدراسة خصائص الأنهار أو تدرجات الجبال. فإذا أدت الملاحظات القائمة على هذا النموذج إلى الوصول إلى استنتاجات لم تطبق بنجاح في الحياة العملية أو لم تساعد على إجراء مزيد من البحوث، فإن هذا النموذج يستغنى عنه ليحل محله نموذج أفضل. وحيث أنه من المستحسن من الناحية الإحصائية أن نستخدم مقاييس الوحدات المتساوية كلما أمكن ذلك (من الواضح أن معظم المقاييس النفسية ليس لها صفر ثابت وبالتالي لا يمكن اعتبارها مقاييس نسبة)، فإننا نبدأ عادة بافتراض أن المقياس الجديد المستخدم إنما هو من مقاييس الوحدات المتساوية. فإذا كان البحث سوف يؤدي إلى نتائج غير مقبولة، فإنه يمكننا أن نعيد تحليل النتائج بأساليب إحصائية تقوم على افتراض أن المقياس

الذي يطبق هنا هو فقط مقياس ترتيب. ففي المثال السابق يحتمل أن يكون الارتباط بين القلق والإبداع أعلى قليلاً على مقياس الوحدات المتساوية مما لو كان المقياس مقياس ترتيب، ولكن الاستنتاج حول مدى دلالة العلاقة سيكون على الأرجح متشابهاً في الحالتين. وبناءً على فكرة النماذج المجردة، رياضية كانت أم غير ذلك، تمكن الإخصائيون النفسيون من أن يتحرروا من إعطاء تفسيرات لا يمكن البرهنة عليها للنتائج التي يحصلون عليها.

ولقد أصبح واضحاً أيضاً ـ كما هو الحال في استخدام اللغة وقواعد التعامل الاجتماعي ـ أن نظم القياس تقوم إلى حد كبير على العُرْف. وطالما أن الباحثين يتبعون نفس الأساليب المتعارف عليها بينهم، فإنهم سوف يفهمون بعضهم بعضاً. وكلما تزايدت البراهين وتراكمت، تغير بعض هذا العُرْف، مثل ما تقوم الهيئات المشرفة على لعبة كرة القدم بتغيير قواعد اللعبة من سنة إلى أخرى.

وبالنسبة لمن يستخدم المقاييس النفسية، فإن أهم احتياط يجب أن يراعيه هو أن يتذكر أنه ليست جميع العمليات التي نعرف كيف نعالج بها الأرقام يمكن أن تطبق بصورة صحيحة على أرقام الاختبارات أو أي تقديرات رقمية أخرى للخصائص النفسية. فنسبة الذكاء التي نجدها مسجلة في السجل المدرسي للطالب إنما هي رقم من نوع مختلف عن مقاييس طوله ووزنه الموجودة قبلها في السجل. والدرجة المئينية التي سجلت لنفس الطالب في اختبارات القراءة والحساب تمثل نوعاً من الأرقام مختلفاً عن كل من نسبة الذكاء والمقاييس البدنية.

وبناء على هذا الأساس من الفهم بأن الأرقام ليست تعني دائمًا ما نظن أنها تعني، فلنتجه الآن إلى بعض الأفكار الإحصائية التي يجب دراستها بإمعان إذا أردنا أن نستخدم المقاييس النفسية بذكاء.

الفصل الثاني

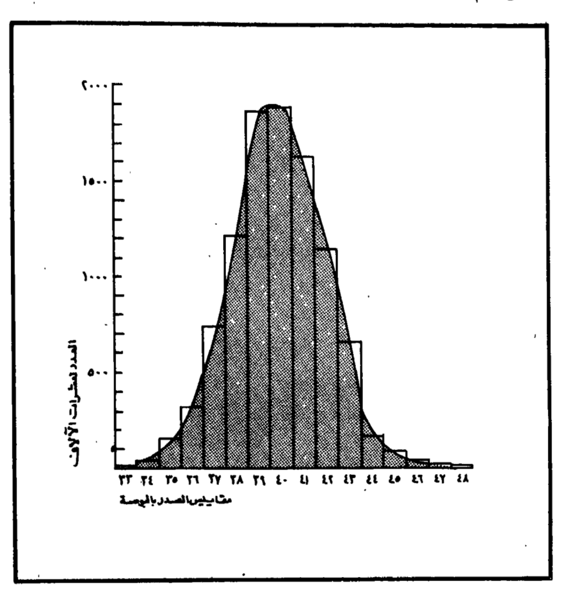
مبادئ إحصائية أساسية

نظام في الاختلاف

في بداية الفصل السابق أثرنا موضوع الاختلاف بين الأفراد. فمنذ أكثر من قرن مضى اكتشف كيتليه Quetelet، وهو رياضي بلجيكي، نظاماً في الاختلاف بين الأفراد حيث تأثر به كثيراً. وكان كيتليه يدرس أساساً بيانات الإحصاء العام ومقاييس الخصائص الفيزيائية. ولم يمض وقت طويل حتى استخدم السير فرانسيس جالتون Francis Galton العالم الإنجليزي النابه طرق كيتليه وطرقاً أخرى اكتشفها بنفسه لقياس بعض الخصائص مثل «حدة الإبصار» ـ وهي سمات نقوم الآن بتصنيفها كخصائص نفسية.

والطريقة التي استخدمها هؤلاء الرواد الأوائل، والتي ما زالت تستخدم حتى الآن، تتلخص في الحصول على قياسات من مجموعة كبيرة من الأفراد، ثم ترتب النتائج من الأدنى درجة إلى الأعلى فيما سموه هم بالتوزيع التكراري. وفي هذه التوزيعات وُجد نفس النموذج مرة بعد أخرى كما يتضح ذلك في دراسة

جالتون لمقاييس محيط الصدر لأكثر من خمسة آلاف جندي، كما هو مبين في الشكل رقم ٢.



شكل ٢ ـ مقاييس الصدر لـ ٥٧٣٨ جندياً. (From F. Galton: **Hereditary Genius: An Inquiry into its Laws**, Appeleton, 1870; reprinted in New York: Horizon Press, 1952, Ch 3)

وعندما يمثل هذا التوزيع في خط بياني يتخذ النموذج شكل منحنى على هيئة الجرس. وعند وسط التوزيع تتجمع مجموعة كبيرة من الحالات مكونة المتوسط أو النزعة المركزية. وبعيداً عن هذا المتوسط نجد نسبة صغيرة من المجموعة الكلية. وعلى ذلك، فهناك عدد كبير من مقاييس الصدر ذات ٣٩، ٤٠، ٤١ بوصة، ولكن هناك عدداً قليلاً فقط من مقاييس الصدر الصغيرة مثل

٣٣ بوصة، أو المقاييس الكبيرة مثل ٤٨ بوصة.

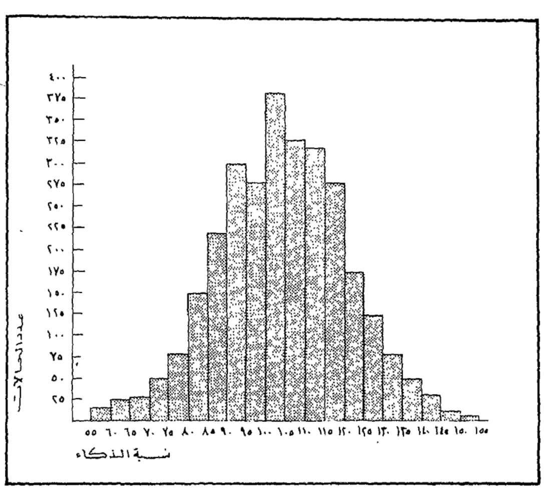
وقد وجد كيتليه وجالتون وآخرون من رواد القرن التاسع عشر هذا الانتظام البديع في توزيع سمة بعد أخرى، مما جعلهم يعتقدون أن هناك قانوناً كونياً يحكم الاختلافات بين البشر.

إن هذا المنحنى المتماثل الجرسي الشكل كان معروفاً للرياضيين. فقد اكتشفوه عند دراستهم لنتائج ألعاب الصدفة - مثل رمي النرد وقطع النقود وما إلى ذلك - واستنتجوا المعادلة الرياضية لمنحنى التوزيع الذي نحصل عليه مع عدد كبير بشكل لامُتناهٍ من عدد الأحداث، حيث يعين كل واحد منها بما نسميه «الصدفة». ومنحنى الاحتمالات الاعتدالي هذا - (أو منحنى جاوس كما يسمى تكريماً للعالم الرياضي الذي وضع معادلته) أصبح أحد الأسس لعلم الإحصاء الحديث. وقد طبقت هذه المعادلة جيداً على التوزيعات المختلفة للمقاييس التي جمعها كيتليه وجالتون. وتستخدم كلمة «اعتدالي» في الإحصاء لوصف هذا المنحنى، ولكنها لا ترتبط بمعنى «صحيح» أو «أحسن».

وبمرور الزمن أصبح من الواضح أن المقاييس الإنسانية وبخاصة مقاييس السمات النفسية لا تعطي دائماً هذا التوزيع الاعتدالي (منحنى جاوس). فبعض هذه التوزيعات يكون ملتوياً بمعنى أن يكون هناك الكثير من الدرجات العالية أو الكثير من الدرجات المنخفضة في المجموعة بحيث لا نحصل على التماثل الموجود في منحنى التوزيع الاعتدالي. وبعض التوزيعات الأخرى تكون مدببة أو مفرطحة بحيث لا تنطبق عليها معادلة المنحنى الاعتدالي. وتتطلب هذه التوزيعات نوعاً خاصاً من المعادلات. ولكن حيث أن الطرق الإحصائية التي تقوم على أساس التوزيعات الاعتدالية تكون أسهل في الاستخدام وأيسر في التفسير من تلك الطرق التي تقوم على أسس رياضية أقل انتشاراً، فإن مؤلفي الاختبارات النفسية غالباً ما يبذلون جهداً خاصاً لجمع البنود التي تعطي توزيعاً اعتدالياً عندما تطبق على المجموعة المراد اختبارها. وعليه فإنه إذا لاحظنا في المجموعة التي سوف نأخذ عنها المعايير الأولية وجود عدد كبير من الدرجات المنخفضة، فإنه يمكننا أن نستبعد بعض الأسئلة الصعبة، وأن نضع محلها أسئلة أكثر سهولة. وإذا لم يكن الانتشار كافياً، فإنه يمكن أن نضيف بعض البنود لنحصل على عدد أكبر من الدرجات الأعلى والدرجات الأدنى. ويجب أن نشير إلى أنه في بعض الأحيان ولأغراض اختبارية يتعمد الدارسون البحث عن

توزيعات غير اعتدالية. وبالاختصار فإن الطرق المستخدمة لتحليل كل سؤال في أي اختبار تجعل من الممكن الآن اختيار البنود التي تعطي أي توزيع مطلوب. علمًا بأن غالبية الاختبارات جيدة التقنين تعطي توزيعات اعتدالية عند تطبيقها على مجموعات كبيرة.

نرى في الشكل رقم ٣ مثل هذا التوزيع لاختبار ظل لسنوات الاختبار الأول في قياس ذكاء الأطفال. وعلى الرغم من وجود قليل من المظاهر البسيطة لعدم الانتظام إلا أنها ليست بالقدر الذي يبعد هذا التوزيع كثيراً عن الشكل الاعتدالي. فهناك عدد كبير من نسب الذكاء حول ١٠٠، وقليل جداً عند الطرفين بالقرب من ٥٥، ١٥٥. (نسبة قليلة جداً من هذه المجموعة حصلت على درجات أقل أو أعلى من نقطتي النهاية التحكميتين المبينتين هنا في الشكل).



شكل ٣ ـ توزيع نسب الذكاء التي حصل عليها بتطبيق الصورة (ل) من اختبار ستانفورد ـ بينيه للأعمار من $2\frac{1}{2}$ إلى ١٨ سنة.

(From Q. Mc Nemar, **The Revision of the Stanford – Binet Scale.** Boston: Houghton Mifflin Company, 1942)

مقاييس النزعة المركزية

عندما نلخص وصفاً لأي توزيع من توزيعات المقاييس لمجموعة ما، فإنه يجب علينا دائمًا أن نصف خاصتين. فأولاً، يجب أن نصف على نحو ما **النزعة المركزية**، وذلك يكون غالباً بحساب ما نسميه عادة بالمتوسط، وهو ما يسميه الإحصائيون **بالمتوسط الحسابي**. وللحصول على قيمة المتوسط الحسابي فإننا نجمع جميع الدرجات ونقسم الناتج على عدد الحالات. وهناك نوعان آخران للمتوسط يفضل الحصول عليهما في بعض التوزيعات، وهما الوسيط والمنوال. **الوسيط** هو الدرجة التي **تتوسط** التوزيع، أو العدد الذي يمثل نقطة بين العددين الأوسطين. أما **المنوال** فإنه الدرجة **الأكثر تكراراً**، أو بمعنى آخر الدرجة التي حصل عليها معظم أفراد المجموعة.

في الجدول رقم (١) يمكن أن نرى المتوسط والوسيط والمنوال لمجموعة من الأرقام. كما نرى أيضاً لماذا نفضل أحد هذه الأرقام على الرقمين الآخرين. فالأبُ الذي يريد أن يعطي ابنه مصروفاً أسبوعياً **يتفق** مع النزعة المركزية لهذه المجموعة سوف لا يكون مقتنعاً بالمتوسط الحسابي كمتوسط في هذه الحالة. ذلك لأن الولدين في هذه المجموعة اللذين يحصلان على دولار واحد كمصروف لكل منهما ليسا نموذجين يمثلان المجموعة على الإطلاق. وإن جمع هذين الرقمين مع بقية الأرقام يؤدي إلى تضخم المجموع، ويصبح المتوسط وهو **٠,٢٩** دولار كبيراً بحيث لا يكون مميزاً لهذه المجموعة ككل. أما الوسيط وهو **٠,٢٢٥** دولار يبدو **معقولاً في هذه الحالة. ولكن الدليل على النزعة المركزية الذي يعتبر أكثر مناسبة لهذه الحالة هو المنوال وهو ٠,٢٥ دولار، فهو المبلغ الذي يتلقاه معظم أفراد المجموعة.**

وإذا كان التوزيع اعتدالياً تماماً فإن المتوسط والوسيط والمنوال ينطبق كل منها على الآخر، بحيث يمكن أن نستخدم أياً منها كمقياس للنزعة المركزية. وحيث أن قليلاً من التوزيعات يكون اعتدالياً بصورة دقيقة، فإنه يجب علينا عادة أن نقرر أي هذه الأرقام (المتوسط والوسيط والمنوال) يكون أكثر مناسبة وأكثرها دلالة. وغالباً ما يتوقف الاختيار على النواحي الأخرى التي سوف نستخدم فيها الأرقام بجانب ذكر المتوسط فقط.

جدول ١ ـ حساب النزعة المركزية لتوزيع المصروف الأسبوعي لعشرين تلميذاً في فصل دراسي.

الولد	المصروف بالدولار	الولد	المصروف بالدولار
١	٠,١	١١	٠,٢٥
٢	٠,١	١٢	٠,٢٥
٣	٠,١	١٣	٠,٢٥
٤	٠,١٥	١٤	٠,٢٥
٥	٠,١٥	١٥	٠,٢٥
٦	٠,١٥	١٦	٠,٣٠
٧	٠,٢٠	١٧	٠,٣٠
٨	٠,٢٠	١٨	٠,٣٥
٩	٠,٢٠	١٩	١,٠٠
١٠	٠,٢٠	٢٠	١,٠٠

المتوسط = ٥,٧٥ ÷ ٢٠ = ٠,٢٩ دولار.
الوسيط = بين ٠,٢٠، ٠,٢٥ أي ٠,٢٣ دولار.
المنوال = ٠,٢٥ دولار.

مقاييس التشتت

والدليل الثاني الذي نحتاج إليه لوصف توزيع الدرجات، هو مقياس ما للتشتت. إلى أي مدى تنتشر الدرجات في التوزيع؟ فمن الواضح أنه من المهم للمعلم ـ على سبيل المثال ـ أن يعرف شيئاً عن المدى الخاص بتوزيع درجات القراءة في الفصل الدراسي. فهل هو ـ أي المدى ـ من ٥٠ إلى ١٥٠، أو من ٩٠ إلى ١١٠. والدرجة المتوسطة في الحالتين هي ١٠٠ تقريباً، ولكن المجموعة في الحالة الأولى تحتاج تنوعاً في مواد القراءة أكثر مما تحتاج في الحالة الثانية. وأسهل الطرق لوصف هذا التشتت هي أن نطرح أقل درجة من أعلى درجة. وهذا يعطي ما يسمى **المدى**. ففي هذا المثال السابق نجد أن المدى في الحالة الأولى ١٠٠ درجة، وفي الثانية ٢٠ درجة.

ومن سلبيات استخدام المدى أنه اذا كانت هناك درجة واحدة عالية جداً أو منخفضة جداً فإن ذلك يؤثر تأثيراً كبيراً. فإذا كانت هناك مجموعة من ١٠٠٠ طفل، على سبيل المثال، فقد نجد طفلاً واحداً حصل على الدرجة ١٠٠ في اختبار الحساب، وإن أعلى درجة تالية هي ٨٢. فإذا قلنا أن المدى هو من صفر

إلى ١٠٠، فإن قولنا هذا يكون مضللاً بعض الشيء، لأن مثل هذا المدى يعطي انطباعاً خاطئاً عن حقيقة التشتت في هذه المجموعة. وهناك دليل أفضل على التشتت في أية مجموعة لا يتأثر كثيراً بالدرجات المتطرفة، وهو **الانحراف المعياري**، والقانون كما يلي:

$$ع = \frac{\sqrt{\text{مج ح}^2}}{ن}$$

(حيث ح هي انحراف الدرجة عن المتوسط، ن عدد الحالات أو الدرجات، مج تعني المجموع، وهي تعني في هذه الحالة مجموع مربعات انحرافات الدرجات عن المتوسط).

جدول ٢ - حساب الانحراف المعياري لتوزيع درجات أحد عشر طفلاً في اختبار للقراءة.

الطفل	الدرجة	الانحراف عن المتوسط	مربع الانحرافات
١	٢٠	+ ٦	٣٦
٢	١٨	+ ٤	١٦
٣	١٥	+ ١	١
٤	١٥	+ ١	١
٥	١٤	٠	٠
٦	١٤	٠	٠
٧	١٤	٠	٠
٨	١٣	- ١	١
٩	١٣	- ١	١
١٠	١٠	- ٤	١٦
١١	٨	- ٦	٣٦
	١٥٤		١٠٨

المتوسط = ١٥٤ ÷ ١١ = ١٤.

التباين = ١٠٨ ÷ ١١ = ٩،٨٢.

الانحراف المعياري = $\sqrt{٩،٨٢}$ = ٣،١.

والجدول رقم ٢ يوضح كيفية حساب الانحراف المعياري. إن الخطوة الأولى هي حساب المتوسط بالطريقة العادية، ثم نطرح قيمة المتوسط من كل

درجة من هذه الدرجات للحصول على الانحراف عن المتوسط. ثم نربع هذه الانحرافات. ثم نقسم بعد ذلك مجموع مربعات الانحرافات على عدد الحالات لنحصل على «متوسط» هذه المربعات، وهو ما يسميه الإحصائيون «تباين» التوزيع. وللحصول على **الانحراف المعياري** نحسب الجذر التربيعي للتباين.

ومن الواضح أن الانحراف المعياري كدليل للتشتت ليس فيه النقص الذي لاحظناه في المدى، فقيمة الانحراف المعياري لا تعتمد كثيراً على الدرجات المتطرفة الوحيدة. بل على العكس، فإن كل درجة في التوزيع تسهم في قيمة الانحراف المعياري. ففي التوزيعات التي تتجمع فيها الدرجات حول المتوسط يكون الانحراف المعياري صغيراً. أما في التوزيعات التي تنتشر فيها الدرجات كثيراً على جانبي المتوسط، فإن الانحراف المعياري يكون كبيراً.

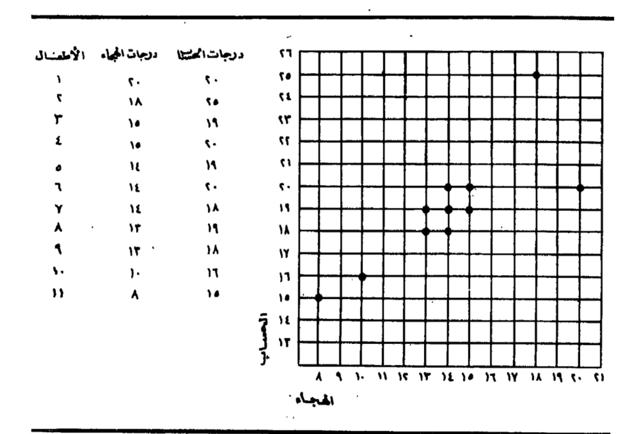
وعندما يكون التوزيع اعتدالياً، فإذا عرفنا قيمة المتوسط والانحراف المعياري فإننا لا نحتاج إلى مراجعة كل التوزيع لمعرفة خصائصه الرئيسية. إذ أنه عن طريق استخدام هاتين القيمتين يمكن باختصار مقارنة الجماعات، وكذلك الأفراد داخل الجماعات. وسوف نتناول كيفية إجراء هذه المقارنات بتفصيل أكثر فيما بعد.

مقاييس العلاقة

بالإضافة إلى مقاييس النزعة المركزية والتشتت، هناك أداة إحصائية أخرى لا يمكن الاستغناء عنها عندما نتناول مقاييس السمات الإنسانية. فعندما يكون من الضروري، كما يحدث في كثير من الحالات، أن نجد **العلاقة** بين مجموعة من الدرجات ومجموعة أخرى، فإننا نستخدم **معامل الارتباط**. وعلى الرغم من وجود طرق عديدة لحساب الارتباط، إلا أن هناك طريقة شائعة الاستعمال يمكن عن طريقها الحصول على دليل يسمى **معامل حاصل ضرب العزوم**.

ومن الممكن كذلك ـ ومن المفيد ـ أن نبين عن طريق الرسم البياني مدى العلاقة بين مجموعتين من الدرجات. ولذلك، فقبل الدخول في تفاصيل حساب معامل الارتباط ـ حاصل ضرب العزوم ـ فلننظر إلى الشكل رقم ٤، وهو عبارة عن رسم بياني يوضح الارتباط بين درجات الهجاء للأحد عشر طفلاً المبينة في الجدول رقم ٢ ودرجاتهم في الحساب. ففي هذا الرسم نجد أن نقطة واحدة تمثل درجة الطفل في كلتا المادتين (الهجاء والحساب). فدرجتا الطفل رقم ٢،

على سبيل المثال، تمثلها نقطة واحدة هي ما تقابل الدرجة ٢٥ على المحــور الرأسي (الحساب)، والدرجة ١٨ على المحور الأفقي (الهجاء). ففي رسم من هذا النوع يمكن لنا أن نحكم على مدى العلاقة بين مجموعتين من الدرجات بناء على اقتراب هذه النقاط من خط (قطر) يمكن رسمه من أسفل زاوية اليسار إلى أعلى زاوية اليمين في هذا الشكل. فإذا كانت جميع النقاط تقع على هذا الخط فإن العلاقة بين المجموعتين هي علاقة تامة كاملة. وفي هذه الحالة فإن كل طفل فوق المتوسط في الحساب يكون فوق المتوسط في الهجاء بنفس القدر. وكل طفل تحت المتوسط في الحساب يكون كذلك وبنفس القدر في الهجاء. ولكن ليست هذه هي الحالة بالنسبة للدرجات المبينة في الشكل رقم ٤. حقيقة هناك ميل كبير للدرجات لأن تقع قريباً من الخط القطري، ولكنها لا تقع عليه تماماً. إن الرسم البياني يعطي صورة حية لهذه العلاقة، ولكنها صورة تقريبية. غير أننا كثيراً ما نحتاج إلى فهم أكثر دقة.



الأطفال	درجات الهجاء	درجات الحساب
١	٢٠	٢٠
٢	١٨	٢٥
٣	١٥	١٩
٤	١٥	٢٠
٥	١٤	١٩
٦	١٤	٢٠
٧	١٤	١٨
٨	١٣	١٩
٩	١٣	١٨
١٠	١٠	١٦
١١	٨	١٥

شكل ٤ ـ رسم بياني يبين العلاقة بين درجات القراءة ودرجات الحساب لأحد عشر طفلاً.

فمعـامل الارتبـاط (ر) هو تعبـير رياضي دقيق عن العـلاقة بين مجموعتين من الأرقام، والقانون كما يلي:

$$\text{ر} = \frac{\text{مج (ح س) (ح ص)}}{\text{ن ع س ع ص}}$$

حيث ح س تعني انحراف الدرجة في المجموعة س عن المتوسط.
ح ص تعني انحراف الدرجة في المجموعة ص عن المتوسط.
ع س الانحراف المعياري للمجموعة س.
ع ص الانحراف المعياري للمجموعة ص
ن عدد الحالات.

جدول ٣ ـ حساب معامل الارتباط بين درجات كل من الهجاء والحساب لأحد عشر طفلاً.

الطفل	درجة الهجاء (س)	درجة الحساب (ص)	حس	حص	حاصل ضرب حس × حص
١	٢٠	٢٠	+ ٦	+ ١	+ ٦
٢	١٨	٢٥	+ ٤	+ ٦	+ ٢٤
٣	١٥	١٩	+ ١	٠	٠
٤	١٥	٢٠	+ ١	+ ١	+ ١
٥	١٤	١٩	٠	٠	٠
٦	١٤	٢٠	٠	+ ١	٠
٧	١٤	١٨	٠	- ١	٠
٨	١٣	١٩	- ١	٠	٠
٩	١٣	١٨	- ١	- ١	+ ١
١٠	١٠	١٦	- ٤	- ٣	+ ١٢
١١	٨	١٥	- ٦	- ٤	+ ٢٤
	١٥٤	٢٠٩			+ ٦٨

متوسط الهجاء ١٥٤ ÷ ١١ = ١٤.
متوسط الحساب ٢٠٩ ÷ ١١ = ١٩.
الانحراف المعياري للهجاء = ٣,١ (حُسِبت في الجدول رقم ٢)
الانحراف المعياري للحساب = ٢,٤. (حُسِبت بالطريقة المبينة في الجدول رقم ٢).

$$\text{ر} = \frac{\text{٦٨}}{\text{١١} \times \text{٣,١} \times \text{٢,٤}} = \text{٠,٨٣}$$

والجدول رقم ٣ يوضح كيفية حساب معامل الارتباط. إن الخطوة الأولى هي حساب المتوسط والانحراف المعياري لكل من المجموعتين من الدرجات (يحسبان بالطريقة المبينة في الجدول رقم ٢). ثم نوجد حاصل ضرب انحراف درجة كل فرد عن المتوسط في حالة الهجاء (ح-) × انحراف درجة نفس الفرد عن المتوسط في حالة الحساب (ح-). ثم نحسب المجموع الجبري لهذه المقادير، ويمثل هذا المجموع بسط الكسر في المعادلة السابقة. وأما مقام الكسر فهو عبارة عن حاصل ضرب عدد الأفراد × الانحراف المعياري لدرجات الهجاء × الانحراف المعياري لدرجات الحساب.

فإذا كانت العلاقة تامة وكاملة بين المتغيرين، فإن معامل الارتباط = ١,٠٠. ويكون معامل الارتباط أقل من الواحد الصحيح عندما تكون العلاقة أضعف. وفي مثالنا السابق نجد أن ر = ٠,٨٣ ويمكننا أن نفهم لماذا يمكن أن يعتبر هذا المعامل عالياً، كما يمكن أن نفهم لماذا لم يصل إلى الواحد الصحيح، إذا ما فحصنا الانحرافات في الجدول السابق. فالأطفال فوق المتوسط في اختبار الهجاء، هم أيضاً فوق المتوسط في اختبار الحساب، باستثناء الطفل رقم ٣. والأطفال دون المتوسط في الهجاء، هم أيضاً دون المتوسط في الحساب، باستثناء الطفل رقم ٨. والطفل رقم ١ أعلى من الطفل رقم ٢ في اختبار الهجاء، بينما نجد الطفل رقم ٢ أعلى من الطفل رقم ١ في الحساب. ولكن يمكن القول عموماً بأن كل طفل من هؤلاء الأطفال ناجح في الحساب بقدر نجاحه في الهجاء. وهذا هو بالضبط نوع العلاقة التي يدل عليها مثل هذا القدر من معامل الارتباط العالي (وهو هنا = ٠,٨٣). وكما أن المتوسط والانحراف المعياري وصفان مختصران لتوزيع معين من الدرجات، فإن معامل الارتباط هو كذلك تلخيص للعلاقة بين توزيعين من الدرجات.

إن كثيراً من معاملات الارتباط التي يتعامل معها الإحصائيون النفسيون تكون أقل بكثير من المعامل الذي وضحناه في مقالنا السابق، بل إنها في بعض الأحيان تكون سالبة. ومعامل الارتباط السالب، إذا حدث أن حصلنا عليه في مثالنا السابق، فإنه سوف يدل على أن الدرجات المنخفضة في الهجاء ترتبط بالدرجات الأعلى في الحساب، والعكس صحيح. وإذا كان معامل الارتباط مساوياً للصفر، فإن ذلك يدل على عدم وجود علاقة، وبمعنى آخر فإن الدرجات الأعلى والأدنى في كل من الحساب والهجاء، إنما تحدث بدون نظام محدد.

ومعامل الارتباط المتوسط، أي حوالي ٠,٥٠، يدل على أن الأفراد الذين يحصلون على الدرجة الأعلى على سمة ما يميلون إلى الحصول على درجة أعلى على السمة الأخرى، ولكن توجد استثناءات كثيرة لهذا الاتجاه العام.

وأساليب قياس الارتباط لا يمكن الاستغناء عنها في بناء الاختبارات العقلية المستخدمة. فإذا لم نكن قادرين على تحليل الخصائص التي ترتبط بعضها ببعض، فلن نتمكن من معرفة ماذا يقيس أي اختبار جديد، أو بأي قدر من الدقة يقيس هذه الخاصة أو تلك. وأما عن كيفية استخدام معاملات الارتباط في هذه الأغراض فسوف نتعرض لذلك في الفصل القادم الخاص بالاختبارات.

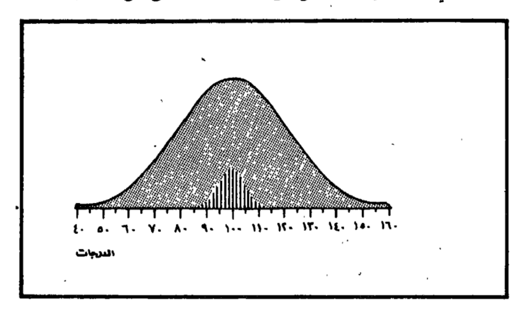
الدلالة الإحصائية والصدفة

كنا نتعامل حتى الآن مع المقاييس الإحصائية التي تصف التوزيعات في جملتها. ولكن هناك أيضاً أساليب إحصائية أخرى تمكننا من الوصول إلى استدلالات عن أحداث أو أشخاص آخرين لم يكونوا أصلاً ضمن المجموعة التي أجريت عليها الدراسة. إن الطرق التي تستخدم لهذا الغرض هي في بعض الأحيان معقدة ويصعب فهمها، ولكن من الضروري أن يفهم فكرتها الأساسية كل من يعمل في ميدان المقاييس النفسية.

إن الفكرة الأساسية هي أن أية مجموعة من الأشخاص أو الأشياء نختارها للقياس إنما تمثل عينة من مجتمع أكبر. فعندما نقوم بالبحث فإننا نقصد في الحقيقة أن نصل إلى استنتاجات حول هذا المجتمع العام، وليس حول العينة التي نطبق عليها الاختبار. وعلى ذلك، فإن الأحد عشر طفلاً الذين درسنا درجاتهم في الهجاء والحساب إنما يمثلون جزءاً صغيراً من مجموعة كبيرة من الأطفال من نفس العمر. وسواء كان مجتمع الأطفال في سن العاشرة الذين نريد أن ندرسهم في بلدة واحدة، أو في محافظة واحدة، أو في دولة واحدة أو في العالم أجمع، فإن الفكرة لا تتغير: هؤلاء الأطفال يمثلون عينة، ونحن نريد أن نستنتج من درجاتهم شيئاً عن الأداء في الهجاء والحساب بالنسبة للمجتمع الذي تنتمي إليه هذه العينة.

إن مشكلة علاقة العينات بالمجتمعات الأصلية قد درست عن طريق الوسائل العملية وعن طريق التفكير الرياضي. فأما عن نوع الدراسة العملية التجريبية، فيمكن للباحث أن يلقي النرد مرة بعد أخرى، ثم يحسب متوسط كل

عشر رميات، ثم يحلل الفروق بين هذه العينات المكونة من عشر رميات والمأخوذة «بالصدفة» من «المجتمع» الأصلي المكوّن من عشرة آلاف رمية على سبيل المثال. إن التفكير الرياضي يحقق نفس الغرض، ولكن بجهد أقل. إن ما سوف نجده في كلتا الحالتين هو أنه عندما نأخذ عينات متتالية من نفس المجتمع فإننا نحصل على توزيع لأي مقياس إحصائي نقوم بحسابه، ويكون هذا التوزيع اعتدالياً إذا كانت العينات مأخوذة من مجتمع موزع توزيعاً اعتدالياً. ومن الممكن تقدير المتوسط والانحراف المعياري لتوزيع المعاينة[1] هذا من المعلومات التي لدينا من عدد قليل من العينات، أو حتى من عينة واحدة فقط.



شكل ٥ ـ توزيع اعتدالي افتراضي للدرجات في مجتمع لانهائي الحجم متوسطه ١٠٠ وانحرافه المعياري ٢٠. وكذلك التوزيع الافتراضي للمتوسطات التي نحصل عليها إذا أخذت عينات كثيرة مكونة من (٢٥) بصورة عشوائية من هذا المجتمع.

(١) «توزيع المعاينة» هو توزيع المقاييس الإحصائية لعدد كبير من العينات التي تؤخذ بالصدفة من مجتمع أصلي واحد. لنفرض، مثلاً، أننا أخذنا خمسين طفلاً في سن العاشرة، ثم حسبنا المتوسط الحسابي لدرجات هؤلاء الأطفال في اختبار للذكاء. ثم أفرض أننا أخذنا خمسين طفلاً آخرين من نفس المجتمع الأصلي، وحسبنا المتوسط الحسابي لدرجاتهم في نفس اختبار الذكاء. فإذا كررنا هذه العملية، وأخذنا عدداً كبيراً من العينات التي يتكون كل منها من خمسين طفلاً من نفس السن، ومن نفس المجتمع الأصلي، وحسبنا المتوسط الحسابي لدرجات أفراد هذه العينات في نفس اختبار الذكاء، لوجدنا أن هذه المتوسطات تتوزع توزيعاً اعتدالياً. إن توزيع متوسطات هذه العينات التي تؤخذ بالصدفة من مجتمع أصلي واحد يسمى «توزيع المعاينة» للمتوسط الحسابي the sampling distribution of the mean. ومن الممكن أيضاً الحصول على توزيع المعاينة للمقاييس الإحصائية الأخرى بنفس الطريقة التي حصلنا بها على توزيع المعاينة للمتوسط الحسابي (المترجم).

في الشكل رقم ٥ نرى توزيع متوسطات العينات المكونة من ٢٥ والمأخوذة من مجتمع متوسطة ١٠٠ وانحرافه المعياري ٢٠. ولا داعي للخوض في العلاقات الرياضية أو القوانين المستخدمة لإجراء هذه الحسابات. ولكن الحقيقة الهامة التي ندركها من الشكل رقم ٥ هي أن تشتت **متوسطات** العينات المكونة من (٢٥) **أقل بكثير** من تشتت درجات المجتمع الأصلي الذي أخذت منه هذه العينات.

وهناك تسمية خاصة للانحراف المعياري لتوزيع المعاينة، حيث أن هذا الانحراف المعياري يدل على مدى تشتت **المقياس الإحصائي** أكثر مما يدل على تشتت الأفراد أو الأحـداث. إنه يسمى **الخطأ المعياري**. وحيث أنه يمكننا أن نحسب أي مقياس إحصائي نشاء سواء في العينات أو في المجتمعات التي تنتمي إليها هذه العينات، فإن هناك أنواعاً عديدة من الخطأ المعياري: الخطأ المعياري للمتوسط، والخطأ المعياري للفرق بين متوسطين، والخطأ المعياري لمعامل الارتباط، والخطأ المعياري للدرجة. وبالتالي فإن الخطأ المعياري، أياً كان نوعه، يوضح لنا مدى الاختلاف الذي نتوقعه في كل العينات المختلفة المأخوذة من المجتمع الأصلي بالنسبة للمقياس الإحصائي المستخدم.

إن مفهوم توزيع المعاينة يساعدنا على تقدير احتمال أن أي نتيجة حصلنا عليها قد جاءت بطريق «الصدفة». والحقيقة أن «الصدفة» تعني الإشارة إلى أي عامل من العوامل غير المعروفة والتي تسبب اختلاف العينات بعضها عن بعض. فلماذا تختلف الأرقام التي نحصل عليها عندما نلقي بالنرد مرات عديدة؟ إنها الصدفة. ولماذا تختلف الإجابات من جماعة إلى أخرى في عملية قياس الرأي العام، حتى ولو تم اختيار هذه الجماعات بنفس الطريقة تماماً؟ إنها الصدفة.

ولتوضيح ذلك فلنأخذ مثالاً نقارن فيه بعض الدرجات حتى نرى عما إذا كان اختلاف المتوسطات هو من عمل الصدفة، أم ان الأمر غير ذلك. لنفرض جدلاً أن معلم الأحد عشر طفلاً المشار إليهم سابقاً أراد أن يستخدم طريقة جديدة على أمل أن يتحسن مستواهم في الهجاء. وبعد شهر من التدريب المكثّف على هذه الطريقة الجديدة، اختبر الاطفال مستخدماً اختباراً يتساوى في الصعوبة مع الاختبار الأول. ويبين الجدول رقم ٤ درجاتهم في الاختبار الأول وفي الاختبار الثاني.

جدول ٤ ـ حساب قيمة «ت» لاختبار الدلالة الإحصائية للزيادة في درجات الهجاء.

الطفل	الدرجة (أ)	الدرجة (ب)	الفرق	الانحراف عن متوسط الفرق (ح)	مربع الانحراف (ح٢)
١	٢٠	٢٣	+ ٣	+ ١	١
٢	١٨	٢٠	+ ٢	٠	٠
٣	١٥	١٩	+ ٤	+ ٢	٤
٤	١٥	١٥	٠	− ٢	٤
٥	١٤	١١	− ٣	− ٥	٢٥
٦	١٤	١٦	+ ٢	٠	٠
٧	١٤	١٨	+ ٤	+ ٢	٤
٨	١٣	١٥	+ ٢	٠	٠
٩	١٣	١٦	+ ٣	+ ١	١
١٠	١٠	١٣	+ ٣	+ ١	١
١١	٨	١٠	+ ٢	٠	٠
	١٥٤	١٧٦	٢٢		٤٠

متوسط الاختبار الأول = ١٥٤ ÷ ١١ = ١٤.

متوسط الاختبار الثاني = ١٧٦ ÷ ١١ = ١٦.

متوسط الفروق = ٢٢ ÷ ١١ = ٢.

الانحراف المعياري للفروق = √(٤٠ / ١٠) = √٤ = ٢.

الخطأ المغياري للفروق = ٢ ÷ √١١ = ٠,٦.

ت = متوسط الفروق / الخطأ المعياري للفروق = ٢ / ٠,٦ = ٣,٣.

دالة > ٠١و (أي دالة عند مستوى أقل من ٠,٠١).

يتضح مــن الجدول السابق أن متوسط درجات المجموعة قد ارتفع من ١٤ إلى ١٦. وفي حين أن هناك طفلاً واحداً حصل على درجة في الاختبار الثاني أقل من درجته في الاختبار الأول، فإن الباقين جميعهم قد تحسن مستـواهم بدرجات متفاوتة. ولكن هل هذا الفرق الذي يعادل درجتين في المتوسط كبير جداً بحيث لا يمكن أن يحدث بمجرد الصدفة إذا أخذنا عينة ثانية من هجاء هؤلاء الأطفال، وبدون إعطائهم أي تدريب خاص في الهجاء؟ للإجابة على هذا

السؤال نستخدم مقياساً إحصائياً معيناً هو ت، وهو عبارة عن نسبة فرق المتوسطين إلى الخطأ المعياري للفروق. (هناك، بالطبع، احتمال آخر، وهو أن زيادة المتوسط بمقدار درجتين لا يمثل اختلافاً راجعاً إلى الصدفة، وإنما يعود إلى تعلم الأطفال من الاختبار الأول. إن أخذ هذا الاحتمال في الاعتبار يحتاج إلى تصميم أكثر تعقيداً. دعنا نهمل هذا الاحتمال في الوقت الحالي). وقبل أن ننظر في معنى ت، دعنا نرى الآن كيف نحسب قيمتها.

الخطوة الأولى هي أن نعين الفرق بين درجتي كل فرد في المجموعة. ثم نحسب بعد ذلك المتوسط والانحراف المعياري لتوزيع هذه الفروق بالطريقة العادية التي نتبعها بالنسبة للدرجات الفردية. ويختلف هذا الحساب من ناحية معينة عن الحساب المبين في الجدول رقم ٢. فنحن في الإحصاء الاستدلالي، على خلاف الإحصاء الوصفي، لأسباب ليس من الضروري أن نذكرها هنا، نقوم بقسمة مجموع مربعات الانحرافات على عدد ينقص واحداً عن مجموع عدد الأفراد في المجموعة، وفي هذه الحالة نقسم على ١٠ بدلاً من ١١ (القانون المستخدم لحساب الانحراف المعياري هنا هو $\sqrt{\frac{\text{مج ح}^{٢}}{\text{ن} - ١}}$). وللحصول على الخطأ المعياري للفرق بين متوسطين (الانحراف المعياري لتوزيع المعاينة لهذه الفروق) نقسم الانحراف المعياري للفروق على الجذر التربيعي لعدد أفراد المجموعة (لسنا في حاجة إلى الدخول في الأسباب الرياضية لهذا الحساب). وعليه فإن قيمة ت هي النسبة بين متوسط الفروق إلى الخطأ المعياري للفروق. وفي هذه الحالة ت سوف تساوي ٣,٣. وبالرجوع إلى الجداول الإحصائية الخاصية بقيم ت، نجد أن هذه القيمة ذات دلالة إحصائية واضحة عند مستوى أقل من ٠١, (وقد عبر عن ذلك في الجدول رقم ٤ بعبارة: دالة < ٠١,). ولتفسير هذه العبارة الأخيرة بلغة إحصائية، نقول إنه ليس أكثر من مرة واحدة في كل مائة مرة نحصل على هذه النتائج عند تطبيق اختبارين في الهجاء على مجموعة من الأطفال مشابهة لمجموعتنا، إذا كانت عوامل الصدفة وحدها هي السبب في الاختلاف. ويمكن، إذن، أن نستنتج، في شيء من الثقة، أن هذه الطريقة الخاصة في التدريس كانت فعالة. وربما كانت أسباب فعاليتها ليست تلك التي يظنها المعلم، وعليه فإنه يصبح من الضروري أن يطبق هذه الطريقة تحت ظروف مختلفة، وأن يقارن بين المجموعات التي تدرس بهذه الطريقة والمجموعات التي تدرس بغيرها، حتى يمكنه أن يتأكد من فعالية هذه الطريقة.

وهناك صيغ متعددة لقانون حساب قيمة ت صممت لتستخدم مع مختلف أنواع البيانات. ولكن في جميع الحالات، فإنّ النتيجة هي تقرير احتمالٍ ما، كما هو واضح في مثالنا السابق. وتقارير الاحتمالات هذه هي ما يشير إليها مفهوم **الدلالة الإحصائية**. فالفرق الدال إحصائياً هو ذلك الفرق الذي يعطي قيمة ت ذات احتمال **قليل**. وهذا يعني أنه من غير المحتمل جداً أن نحصل على هذا الفرق في عينات الصدفة، أي تلك التي لم يتم اختيارها بطريقة خاصة، أو التي لم تتلق معاملة من نوع خاص. ومعامل الارتباط الدال إحصائياً هو ذلك المعامل الذي لا يحتمل الحصول عليه إذا رتبنا الأرقام عشوائياً في أزواج، ثم قمنا بحساب معامل الارتباط. وكلما انخفضت قيمة الاحتمال (أي مستوى الدلالة الإحصائية)، كنا أكثر تأكداً من أن النتائج تمثل شيئاً آخر غير الصدفة.

وكل من يقرأ تقارير البحوث سوف يقابل عبارات مثل «دال عند مستوى ٠,٠٥». وهذه طريقة لتقرير ما سبق أن أشرنا إليه ـ فهي تقول أن مثل هذه النتائج نتوقع ألا نحصل عليها أكثر من خمس مرات في كل مائة مرة عن طريق الصدفة وحدها. وقد جرى العرف على ألا تعتبر النتائج «دالة إحصائياً» إذا كان مستوى الدلالة أكثر من ٠,٠٥، ولو أن مستوى الدلالة قد يختلف بناء على ما يريد أن يصل إليه الباحث.

قد يبدو ذلك أمراً معقداً وخارجاً عن الموضوع بالنسبة للمبتدىء، إلا أنه يمثل المستوى الأولي جداً من الإحصاء الاستدلالي. فالسؤال حول ما إذا كان الفرق الذي يبني عليه الباحث استنتاجاً ما يمكن أن يحدث بالصدفة، إنما هو سؤال هام. وهناك أنواع أخرى من الاستدلالات التي كثيراً ما يحتاج الباحثون إلى القيام بها. فهم أحياناً يحتاجون إلى الحصول على تقدير دقيق بقدر الإمكان لقيمة المتوسط أو الانحراف المعياري للمجتمع الأصلي الذي أخذت منه العينة. وأحياناً يريد الباحثون معرفة مدى كفاءة النموذج الرياضي الذي اقترح ليمثل النتائج التي حصلوا عليها في سلسلة من التجارب. وأحياناً يصبح من الضروري معرفة كم شخصاً يجب أن يضافوا إلى العينة التي تجري عليها التجربة لكي نحصل على إجابة محددة لسؤال معين في بحث ما. ولقد أصبح وضع الأساليب الإحصائية الخاصة لحل هذه المشكلات وكثير من أنواع المشكلات الأخرى يمثل تحدياً مستمراً للمتخصصين في الإحصاء الرياضي. ولكي يكون الإنسان عالماً لا يكفي فقط أن يقوم بتخطيط وإجراء التجارب،

ولكن يجب أيضاً أن يخرج باستنتاجات صحيحة، بل باستنتاجات صحيحة فقط من نتائج هذه التجارب.

إن تطوير الأساليب الإحصائية التي تسمح للباحثين بالقيام باستدلالات على مجتمعات بأكملها عن طريق عينات منها، قد مكننا من أن نزيد معارفنا بصورة أسرع مما لم تكن هذه الأساليب متوفرة. وبالإضافة إلى ذلك، فإن فهم هذه الأساليب، ولو بصورة عامة على الأقل، أمر مفيد لكل مواطن ذكي، بدون أن يكون هو نفسه باحثاً، من حيث أن ذلك سوف يساعده على تقييم ما يسمع وما يقرأ عن الناس وعن السلع التجارية، وعلى القيام ببعض الاستنتاجات حول ما إذا كانت الأرقام التي تقدم كبرهان على تفوق معجون أسنان جديد، مثلاً، تبرهن، في الحقيقة، على أي شيء ما. ومهما كان التفكير المتضمن في الاستدلالات الإحصائية معقداً، فإن معرفته أمر جدير جداً بأن نبذل الجهد من أجله.

الفصل الثالث

الاختبارات النفسية

الاختبارات في مقابل المقاييس

حتى الآن لم ننظر فيما إذا كان هناك أي فرق بين مصطلح **الاختبار** ومصطلح **المقياس**. وعلى الرغم أن هناك تداخلاً في المعنى إلا أنهما ليسا مترادفين تماماً. فمصطلح المقياس يستخدم في ميادين كثيرة من ميادين البحوث النفسية حيث لا يكون مصطلح الاختبار مناسباً للاستخدام. فعلى سبيل المثال، في التجارب التي تدرس الإحساس، والإدراك الحسي، والحكم، استخدم الباحثون كثيراً من **الطرق النفسية الفيزيائية** (السيكوفيزيائية)، بمعنى قياس الكم الفيزيائي المقابل للكم النفسي (مثل درجة وضوح الضوء، ودرجة ارتفاع النغمة). فإذا كان موضوع الدراسة هو البحث عن الحد الأدنى والحد الأعلى لقوة السمع عند الإنسان، فإنهم يقيسون مقدار التردد. وعليه، فإن القياس الفيزيائي يستخدم للإجابة عن سؤال سيكولوجي.

ومن المعتاد أن نقول عن القياس النفسي أنه اختبار إذا كنا نستخدمه أساساً لتقدير بعض خصائص **الفرد** أكثر مما نجيب على سؤال عام، مثل ذلك

السؤال عن قوة السمع عند الإنسان في الفقرة السابقة. ومن الممكن، بالطبع، استخدام مقاييس درجات عتبات السمع كاختبارات بهذه الطريقة. غير أن الاختبار، في الأغلب، عبارة عن أسئلة أو مهام تقدم للشخص المطلوب اختباره، وإن الدرجات التي نحصل عليها لا يعبر عنها بوحدات فيزيائية من أي نوع.

وعلى ذلك، فليست المقاييس اختبارات، وكذلك، فإن العكس صحيح أيضاً، فليست كل الاختبارات مقاييس. فهناك بعض اختبارات الشخصية، مثلاً، لا يحصل المفحوص فيها على درجات. وقد يستخدم الإخصائي النفسي مثل هذا الاختبار لكي يساعده على إعداد وصف لفظي للفرد. وهو هنا ليس في حاجة إلى أن يلجأ إلى القياس في أي مستوى من المستويات التي ميزنا بينها في الفصل الأول. **فالقياس**، كما سبق إن أوضحنا في الفصل الأول، هو عملية استخدام الأرقام في تحديد الأشياء حسب قواعد معينة. و«الاختبار» يمكن أن يعرف على أنه موقف مقنن صمم خصيصاً للحصول على عينة من سلوك الفرد. وعندما يعبر عن هذه العينة بالأرقام، فمن الممكن أن نستخدم كلمة اختبار أو مقياس.

وعلى ذلك، فإنه بالرغم من أن التداخل بين هذين المفهومين ليس كاملاً، فإنه لا يزال في إمكاننا أن نقول أن معظم الاختبارات هي وسائل قياس، وأن معظم المقاييس النفسية يمكن أن تستخدم كاختبارات. وسوف نعالج في هذا الفصل بعض الأفكار التي تنطبق على الاختبارات على وجه خاص.

خلفية تاريخية

منذ أن بدأ علم النفس كعلم - أي منذ إنشاء **ڤلهلم ڤونت** Wilhelm Wundt لمختبره في مدينة لايبزج Leipzig في عام ١٨٧٩ - ركز الباحثون جهودهم ابتداءً على البحث عن مبادىء عامة يمكن أن تطبق على كل فرد. فقد حاول الإخصائيون النفسيون اكتشاف كيفية عمل عين الإنسان ومخه معاً لإدراك اللون والشكل والحجم، وهي إدراكات حسية تحدث لنا جميعاً. كما صمموا التجارب لدراسة عملية التعلم، ومن ثم استنتجوا «قوانين التعلم». كما درسوا أيضاً نمو الإنسان، وذلك بمقارنة مجموعة ذات مرحلة عمرية معينة بمجموعة أخرى ذات مرحلة عمرية أخرى، وقاموا بتحديد معايير السلوك لكل مرحلة عمرية،

ووضعوا نظريات حول عملية النمو.

وإلى جانب ذلك، كان هناك اتجاهاً ثانوياً للاهتمام والجهد. فقد تنبه بعض الباحثين الأوائل في المختبرات الألمانية لعلم النفس التجريبي الحديثة النشأة، إلى الاختلاف الذي يجدونه بين الأفراد المفحوصين. فهؤلاء الباحثون الذين كانت لهم في كثير من الأحيان اتجاهات عملية، أدركوا أن قياس الفروق بين الناس قد يكون له تطبيقات هامة في المدارس، والمصانع، والمكاتب. وقد كان جيمس ماكين كاتل James Mckeen Cattell ـ وهو أمريكي درس في مختبر فونت ـ ذا تأثير قوي في حركة استخدام المقاييس النفسية كاختبارات عقلية. وقد كان هو، في الحقيقة، أول من استخدم لفظ «الاختبارات العقلية» في عام ١٨٩٠، ولكن كان هناك آخرون ظهروا بوضوح في هذه الحركة خلال السنوات الأخيرة من القرن التاسع عشر.

إن ما كان يتطلع إليه هؤلاء العلماء كثيراً هو إيجاد طريقة كمية لتقدير الذكاء العام. وكانوا يعتقدون أنه يمكنهم الحصول على دليل للذكاء، إذا أمكنهم قياس جميع الخصائص وهي متجمعة معاً في الأفراد، تلك التي كان يقاس كل منها على حدة في المختبرات التجريبية ـ الإحساس، والإدراك الحسي، والانتباه، والتمييز، وسرعة رد الفعل، وهكذا. والإنسان المتفوق ـ حسب تفكيرهم هذا ـ هو الشخص الذي يحصل على درجات عالية في كل هذه الخصائص. ورغم أنه ليس هناك خطأ واضح في هذا التفكير، إلا أن محاولة قياس الذكاء بهذه الطريقة لم تنجح، ذلك لأنه عندما قام الإخصائيون النفسيون بتحليل مقاييسهم اكتشفوا أن هذه السمات لم ترتبط بعضها ببعض ارتباطاً وثيقاً. وفضلاً عن ذلك، فإن مجموع الدرجات لم يبد أنه من الممكن اعتباره دليلاً على الذكاء العام. فالتلاميذ الضعفاء، على سبيل المثال، حصلوا على درجات عالية شأنهم في ذلك شأن التلاميذ المتفوقين. لقد أصبح واضحاً أن مشكلة قياس الذكاء كانت في حاجة إلى أسلوب مختلف.

وسرعان ما حدث هذا الانطلاق الجديد في فرنسا، عندما بدأ ألفرد بينيه Alfred Binet بافتراض أن الذكاء خاصة مركبة فطرية، وليس مجرد مجموع سمات بسيطة كثيرة. ولقياس الذكاء، من وجهة نظر بينيه، لا بد من إيجاد وسائل لتقييم كيفية تعامل الأفراد مع المواقف التي تتطلب التفكير، والحكم، وحل المشكلات. وعلى مدى سنوات عديدة قام بينيه بتجريب أنواع كثيرة من

المهام كاختبارات محتملة لذكاء الأطفال على أطفاله وعلى أطفـال من مختلف الأعمار في المدارس الفرنسية في باريس. وقام أخيراً في عام ١٩٠٥، بالتعاون مع تيوفيل سيمون Theophile Simon، بنشر أول اختبار حقيقي للذكاء، والذي يُعْتَبر الأصل بالنسبة لكل الاختبارات الحالية.

وبعد ذلك حدث تقدم سريع. فقد عدل اختبار بينيه وسيمون ليستخدم في كثير من الدول، كما صممت اختبارات للكبار بالإضافة إلى الاختبارات التي صممت خاصة للأطفال. وكذلك وضعت الاختبارات الجمعية أثنـاء الحرب العالمية الأولى، ثم طورت سريعاً بعد ذلك لتستخدم في المدارس والمؤسسات الصناعية. وصممت كذلك اختبارات لكثير من القدرات التي ليست في مثل شمول وعمومية الذكاء. وأخذت محاولات قياس خصائص الشخصية وكذلك القـدرات تزداد انتشـاراً. وفي منتصف القـرن العشـرين تـوفـرت آلاف من الاختبارات على اختلاف أنواعها، منها الجيد، ومنها الرديء، ومنها مـا هو وسط.

وفي أثناء كل هذا النشاط في ميدان الاختبارات العقلية، ظهرت المعايير والمبادىء. وهذه تستخدم الآن كإرشادات مفيدة لمن يريد أن يبني اختبارات جديدة، أو لمن يحتاج أن يختار بحكمة من الاختبارات الموجودة. وفضلاً عن ذلك، فقد أصبـح من الضروري، في عـالمنا الحـديث الذي يشعـر بأهميـة الاختبارات، أن يعرف كل رجل مثقف أو امرأة مثقفة شيئاً عن هذه المعايير والمبادىء. فالمدرس الذي يريد أن يعدّل طريقته التدريسية لتلائم الخصائص المتباينة لتلاميذه، والأب الذي يرغب في مسـاعدة ابنـه أو ابنته عـلى القيام بتخطيط حكيم يتعلق بالدراسة أو المهنة، ورجل الأعمال الذي لا يريد أن يخدعه بعض من لا خُلق لهم من ناشري الاختبارات وبائعيها، كل هؤلاء في حاجة لمعرفة عملية عن كيفية بناء الاختبارات، وعن كيف يمكن الحكم عليها.

معنى صدق الاختبار

إن أهم ما يؤخذ في الاعتبار هو الصدق الذي يتصل مباشرة بالسؤال الذي يقول: «ماذا يقيس الاختبار؟» وإذا لم يكن لدينا إجابة كافية بـدرجة معقولة لهذا السؤال، فإن أي اختبار سوف يكون عديم الفائدة في محاولاتنا للتعامل بحكمة مع البشر، صغار أو كبار، نحن أو الآخرين. بل ربما كان الموقف أكثر سوءاً من عدم الفائدة، لأننا لو اعتمدنا على الافتراض الخاطىء فيما

يتعلق بمعنى درجات الفرد، فقد نوجهه إلى قرارات قد تؤدي إلى عدم التوافق والوقوع في أخطاء جسيمة.

وهناك كثير من الأشخاص لا يستطيعون معرفة أن عنوان الاختبار لا يقول لنا في الحقيقة شيئاً على الإطلاق عما يقيسه هذا الاختبار. فكل إنسان يستطيع أن يكتب مجموعة من الأسئلة، ويكون واثقاً من أن هذه الأسئلة تدور حول عملية «التفكير» أو «الاستعداد الميكانيكي» أو «المرونة المعرفية». ولكن معرفة أي العمليات العقلية يجب أن يستخدمها المفحوص في الواقع للإجابة على هذه الأسئلة، فهي مهمة شاقة وطويلة. فقد يبني شخص ما اختباراً لقياس «القدرة على التفكير الاستدلالي»، وربما يثبت أن هذا الاختبار إنما يقيس اتجاهات الطبقة المتوسطة. وقد يحدث أن اختباراً في «الاستعداد الميكانيكي» قد يقيس فعلاً الذكاء العام، وأن اختباراً في «النضج الانفعالي» قد يقيس فقط ما يعرفه المفحوصون عن التقبل الاجتماعي لنمط معين من السلوك. إن المهارة الأولى التي يجب أن يجيدها الفرد عند تقييم صدق الاختبارات هي عدم أخذ عناوين الاختبارات في الاعتبار. فالأمر المهم هو ليس **التسمية** التي يضعها المؤلف لهذا الاختبار، ولكن بماذا **تتعلق** درجات هذا الاختبار.

وفي بداية حركة القياس كانت الطريقة المقبولة هي أن يعرّف الفرد ماذا يريد أن يقيس، ثم يجمع بعد ذلك الأدلة والبراهين لإثبات مدى نجاح الاختبار في قياس ما أراده الفرد. وخلال هذه الفترة، كما رأينا سابقاً، كان معظم الجهد موجهاً لقياس الذكاء العام. وقد كان واضحاً ، أنه على الرغم من تزايد عدد الباحثين، إلا أنهم لم يتفقوا جميعاً على تعريف محدد لهذا المفهوم المحيّر، ولكنهم قد افترضوا جميعاً، من أجل الأغراض العملية، أن الذكاء هو السمة التي يعتمد عليها الحكم بالتفوق أو الغباء في الدراسة، ومن ثم بحثوا في المواقف المدرسية عن البرهان على صدق اختبارات الذكاء. وكان الطريق هو تقدير مقدار العلاقة بين درجة النجاح في الدراسة ـ كما يمكن أن يتم التنبؤ بها من درجات الاختبار ـ ودرجة النجاح الحقيقية. وبناء على الأدلة التي حصلوا عليها، تم استبعاد القياسات المختبرية مثل التمييز، وزمن الرجع، وما إلى ذلك، واستبقيت الأسئلة من نوع اختبار بينيه ذات الاستجابات المركبة لتكون أساساً لاختبارات الذكاء.

وقياسات إحدى السمات النفسية في موقف ما من مواقف الحياة

الحقيقية، مثل الفصل الدراسي، تسمى قياسات المحك. والارتباطات بين درجات الاختبار وقياسات المحك هذه تسمى معاملات الصدق.

فمنذ البداية، إذن، كان الصدق يعني مدى قياس الاختبار لما وضع لقياسه. ولهذا كان يحسب معامل الارتباط بين درجات الاختبار وبين مقاييس المحك لمعرفة صدق الاختبار. ولكن بمضيِّ الوقت اتضح أن هناك تعقيدات في كل من الطريقة والمفهوم. ومما يؤسف له، أنه لم يكن من الممكن في معظم الحالات أن نجد محكاً واحداً يمكن أن يعتبر دليلاً واضحاً على سمة عقلية. فقد يقرر باحثان يدرسان نفس القدرة ـ مثل الاستعداد الميكانيكي ـ استخدام محكات مختلفة، وبالتالي قد يحصل كل منهما على نتيجة تختلف عن الآخر. فالباحث الأول قد يرى أن المحك المناسب هو درجات التلاميذ في مقرر الأعمال (الأشغال) اليدوية. والباحث الثاني قد يرى أن المحك المناسب هو طول الوقت الذي يستغرقه الأفراد الجدد لتعلم مهارة ميكانيكية بسيطة في المصنع. والآن، ماذا نعمل عندما نجد أن معامل الارتباط بين درجات الاختبار والمحك هو في الحالة الأولى ٠,٠٦، وفي الحالة الثانية ٠,٥٩ كيف يتسنى لنا أن نقرر ما إذا كان الاختبار صادقاً إذا أعطانا نتائج من هذا النوع؟ فهل هذا الاختبار يقيس الاستعداد الميكانيكي أم لا؟

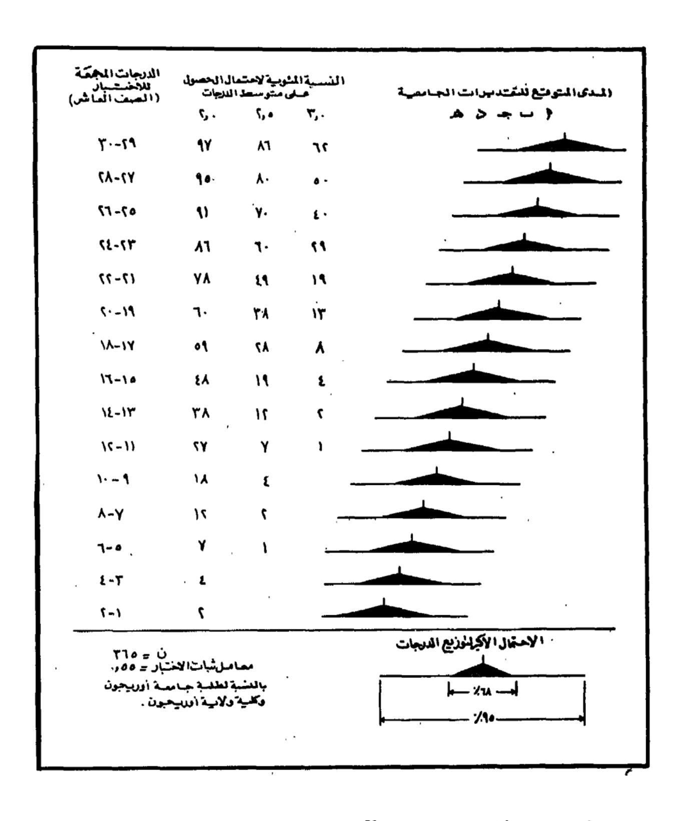
لقد اتضح من واقع الخبرة، كهذه الحالة السابقة وأمثالها، ان عملية التحقق من صدق الاختبار إنما هي عملية طويلة، وهي لا تقوم على حادثة مفردة. ونحن نستطيع فقط أن نفهم ماذا يقيس الاختبار حقاً، إذا ما قمنا بدراسة ارتباطات الاختبار بمجموعة مختلفة من المحكات. وإذا توفرت سلسلة من البحوث حول اختبار «الاستعداد الميكانيكي» ـ على سبيل المثال ـ فقد توضح أنه يقيس في الحقيقة القدرة على القيام بالحركات الماهرة ذات الضبط الدقيق، ولا علاقة له بالقدرة على فهم العلاقات المعقدة بين الأجزاء الميكانيكية. ومن ثم، فقد يرتبط هذا الاختبار ارتباطاً عالياً مع درجات أعمال النجارة، ولكنه لا يرتبط ارتباطاً عالياً مع درجات الأعمال الميكانيكية. وقد يساعد هذا الاختبار على اختيار عمال قادرين لمصنع دون آخر.

ولذلك، فبدلاً من أن نسأل السؤال التقليدي عن الصدق وهو: «إلى أي مدى يقيس هذا الاختبار ما وضع لقياسه؟» فإننا نميل الآن إلى أن نسأل: ماذا يقيس هذا الاختبار فعلاً؟. ونحن نعلم الآن أنه يجب أن نقوم بتحليل محتوى

الاختبار ونفحص كثيراً من معاملات الارتباط بين درجات الاختبار ومجموعة مختلفة من المحكات في جماعات مختلفة قبل أن نصل إلى إجابة لهذا السؤال. ومن خلال هذا العمل فإن معلوماتنا عن السمات العقلية الأساسية، وكيف يرتبط كل منها بالأخرى، إنما تزيد باستمرار وتتغير، كما أن معلوماتنا حول الاختبارات ذاتها تزيد أيضاً وتتغير. فليس من الضروري في الواقع أن يحدد الإخصائي النفسي بدقة في بداية بحثه السمة التي يأمل أن يقيسها الاختبار. فإذا كانت لديه فكرة عامة عن هذه السمة وعلاقتها بالمفاهيم النظرية أو بالمواقف العملية، فإن دقته في تحديد معنى السمة سوف تزيد كلما جرب الاختبار في سلسلة من البحوث المنفصلة.

إن فهم مضمون عملية التحقق من صدق الاختبارات يحمل ثلاثة معانٍ واضحة لمستخدمي الاختبارات. أولها، أنه إذا كان الاختبار سوف يستخدم في اتخاذ قرارات حول الأفراد أو الجماعات، فلا بد من دراسة كل البيانات المتوفرة قبل أي محاولة لتفسير درجات الاختبار. وهذا يعني أن من يستخدم الاختبار يجب أن يكون قادراً على إجراء مثل هذه الدراسة. «فبرنامج تطبيق الاختبارات» في المدارس أو الصناعة يمكن أن يؤدي وظيفته بفعالية في حالة ما إذا أشرف عليه شخص على درجة عالية من القدرة والكفاءة.

والمعنى الثاني هو أنه عندما يستخدم الاختبار من أجل التنبؤ أو الاختيار (مثل اختيار المتقدمين لعمل ما)، فإنه يجب أن يعين صدقه في الموقف النوعي الذي سوف يستخدم فيه، كلما كان ذلك ممكناً. وهذا يتطلب، بطبيعة الحال، دراسة خاصة في نفس الموقف الذي سيستخدم فيه الاختبار، وذلك قبل أن نبدأ باستخدام الاختبار بصورة روتينية في انتقاء المتقدمين. وفي حالة التخطيط لمثل هذه الدراسة فإن السؤال المناسب الذي علينا أن نسأله فيما يتعلق بصدق الاختبار لن يكون: «هل هذا الاختبار صادق؟»، ولكن سوف يكون السؤال: «هل يضيف هذا الاختبار جديداً لصدق الطرق المستخدمة حالياً في عملية الاختيار؟». فقد ترى إحدى الكليات التي تطبق معايير محددة للاختيار على أساس درجات المدرسة الثانوية أن ما يسمى باختبار «الاستعداد الجامعي» لا علاقة له بالبرنامج المطبق في هذه الكلية. وقد ترى كلية أخرى ذات برنامج مختلف، وتستمد طلبتها من نوعيات مختلفة، أن اختبار «الاستعداد الجامعي» ذاته على درجة عالية من الصدق. غير أنه ليست جميع المواقف تسمح بإجراء صدق الاختبار بهذه



الدرجات المجمعة للاختبار (الصف العاشر)	النسبة المئوية لاحتمال الحصول على متوسط الدرجات ٢,٠	٢,٥	٣,٠
٣٠-٢٩	٩٧	٨٦	٦٢
٢٨-٢٧	٩٥	٨٠	٥٠
٢٦-٢٥	٩١	٧٠	٤٠
٢٤-٢٣	٨٦	٦٠	٢٩
٢٢-٢١	٧٨	٤٩	١٩
٢٠-١٩	٦٠	٣٨	١٣
١٨-١٧	٥٩	٢٨	٨
١٦-١٥	٤٨	١٩	٤
١٤-١٣	٣٨	١٢	٢
١٢-١١	٢٧	٧	١
١٠-٩	١٨	٤	
٨-٧	١٢	٢	
٦-٥	٧	١	
٤-٣	٤		
٢-١	٢		

ن = ٣٦٥
معامل ثبات الاختبار = ٠,٥٥
بالنسبة لطلبة جامعة أوريجون وكلية ولاية أوريجون.

جدول ٥ - بعض الأدلة على صدق اختبارات آيوا للنمو التربوي كمقاييس للقدرة على تحصيل الدراسة الجامعية.

(From J. S. Carlson and D. W. Fullmer. College Norms. Eugene, Ore.: University of Oregon Counseling Center. 1959, p.11.)

الطريقة المباشرة السريعة، ولكننا نوصى بها إذا سمح الموقف بذلك. وكمثال لدراسة الصدق التَّنبُّئي لاختبارات آيوا للنمو التربوي[1] والتي بنيت خصيصاً لمعلمي المرحلة الثانوية، انظر إلى الجدول رقم ٥. يبين الجدول، مثلاً، أن طلبة الصف العاشر الذين يحصلون على درجة ٢٥ أو ٢٦ في اختبارات آيوا من المحتمل أن ينجحوا في دراستهم الجامعية. فمن كل ١٠٠ طالب حاصلين على هذه الدرجة (٢٥ أو ٢٦) نجد أن ٩١ طالباً يحصلون على الأقل على معدل تقديرات ٢ (جيد)، و٧٠ طالباً يحصلون على ٢,٥ (جيد +)، و٤٠ طالباً يحصلون على معدل ٣ (أي جيد جداً).

وأخيراً، فإذا كنا نستخدم الاختبارات في المواقف العملية التي نتعامل فيها مع الأفراد، أو في البحوث البحتة التي تهدف إلى زيادة معلوماتنا النظرية عن الفروق الفردية، فإنه يجب أن نتذكر دائمًا أن أفكارنا عن ماهية السمات، وكذلك أفكارنا عن ما تقيسه الاختبارات يجب أن تتعدل إذا أتيح لنا الحصول على أدلة جديدة.

فعندما لا ينجح الباحث في الحصول على معامل الارتباط المتوقع بين «القلق» و «القابلية لتوتر الفشل»، فقد يحتاج إلى تغيير أفكاره عن الشيء الذي يقيسه اختبار القلق الذي استخدمه في بحثه. وفضلاً عن ذلك، فقد يكون عليه أيضاً أن يعدل من نظريته حول كيفية تأثر الشخصية بتهديد الفشل. وعلى الذين يستخدمون الاختبارات الخاصة بكثرة، مثل المعلمين في المدارس العامة، أن يتابعوا عن كثب المعرفة الحديثة عن هذه الاختبارات، وكذلك عن الخصائص التي تقيسها هذه الاختبارات، إذ أن هناك كثيراً من المعلمين في المدارس حالياً، على سبيل المثال، يتخذون القرارات بالنسبة للأطفال بناء على مفاهيم عتيقة للذكاء. والحقيقة أن هؤلاء المعلمين ليسوا في حاجة فقط إلى اختبارات في الذكاء على درجة أكبر من الصدق، ولكنهم يحتاجون كذلك إلى معرفة أكثر دقة لمعنى كلمة الذكاء ـ فنحن لا يمكن في الحقيقة أن نبحث في صدق الاختبار دون أن نختبر في نفس الوقت صدق أفكارنا.

معنى الثبات

إن الصدق هو أهم ما يجب أن يؤخذ في الاعتبار عند بناء الاختبارات

(١) Iowa Tests of Educational Development.

بكافة أنواعها، وكذلـك عند استخدامها. يلي ذلك في الأهمية الثبات. ولتوضيح معنى الثبات، يمكننا أن نبدأ مرة أخرى بسؤال هو: «بأي قدر من الدقة، أو بأي قدر من الاتساق، يقيس هذا الاختبار السمة التي يقيسها، أياً كانت هذه السمة؟.

عندما يؤدي الفرد اختباراً ما فإن درجته على هذا الاختبار قد تتأثر بعوامل عديدة بعيدة كل البعد عن القدرة أو سمة الشخصية التي صمم الاختبار من أجل قياسها. فإذا تشتت انتباه المفحوص أو كان متمرداً فقد لا يعطى إجابات صحيحة للأسئلة مثلما يعطي عندما يكون مركزاً على الاختبار. وبالتالي فإنه يحصل على درجة أقل مما يجب. ومن ناحية أخرى، فإنه لو حدث وشاهد المفحوص ـ في الليلة السابقة للاختبار ـ برنامجاً في التلفزيون أو اشترك في مناقشة عابرة والتقى ببعض الأسئلة الخاصة التي يحتويها الاختبار، فإنه في هذه الحالة يحتمل أن يحصل على درجة أعلى مما يؤهله له أداؤه في الظروف العادية. وهذان مثالان على مدى تأثر درجات الاختبار بعوامل الصدفة. وهناك عوامل أخرى عديدة مثل ردود الفعل الانفعالية بالنسبة للفاحص، ودرجة الحرارة والتهوية في غرفة الاختبار، والحظ الحسن أو السيئ في تخمين إجابات الأسئلة التي لا يعرفها المفحوص، وهكذا. وهذه العوامل لا يمكن التنبؤ بها، وهي لا تؤثر في جميع المفحوصين بنفس الطريقة أو بنفس الدرجة.

إن ما يجب على الإخصائي الذي يبنى الاختبار أن يكتشفه ويوصله إلى الإخصائي الذي يستخدم الاختبار هو إلى أي مدى سوف تتأثر دقة تقـدير درجات الأفراد بعوامل الصدفة هذه. فإذا توفرت هذه المعلومات، استـطاع الإخصائي الذي يفسر درجات الاختبار أن يأخذ ذلك في اعتباره، وسوف تكون خلاصة تقريره عن شخص ما على النحو التالي: الفرد أ ذكاؤه فوق المتوسط بقليل، فهناك احتمال ٥٠ ـ ٥٠ إن نسبة ذكائه تقع بين ١٠٥ و١١٥. ولكنه لن يستنتج: إن الشخص أ له نسبة ذكاء تساوي ١١٠.

وهناك طرق عديدة لتحليل ثبات الاختبار وإعداد نتائج التحليل، ولكن معظمها يبدأ من منطلق واحد: وهو تطبيق صورتين من نفس الاختبار على مجموعة من الأفراد لهم نفس الخصائص العمرية والاجتماعية لمجتمع الأفراد الذي يجهز له الاختبار. فإذا كنا نريد أن نعين درجة ثبات اختبار ذكاء خاص

بأطفال سن الثامنة، فإننا نطبق الصورة أ والصورة ب على مجموعة أطفال في نفس السن وفي نفس الوقت. ولكن إذا لم تتوفر غير صورة واحدة، فإننا نقسمها إلى نصفين - وعادة ما يكون النصف من الأرقام الفردية، والنصف الآخر من الأرقام الزوجية - لكل منهما له مفتاح خاص للتصحيح. وبالتالي فإننا نحصل على درجتين لكل فرد، بدلاً من درجة واحدة. وإذا كنا نريد تعيين درجة ثبات اختبار في مهارة استخدام الأصابع لتلاميذ في مدرسة مهنية، فغالباً ما سوف نطبق الاختبار مرتين على نفس المجموعة من التلاميذ، مع مراعاة وجود فاصل زمني قصير بن التطبيقين. وهذه الطريقة أيضاً تمدّ الفاحص بدرجتين لكل فرد. وفي كل مثال من الأمثلة السابقة، ليس هناك سبب لأن نفترض أن جميع عوامل الصدفة سوف تؤثر على الدرجتين بنفس الطريقة، لذلك، فإنه يمكن أن نقول إنه كلما كانت درجتا كل فرد في التطبيقين أكثر تشابهاً، كان الاختبار أكثر تناسقاً، أو دقة، أو ثباتاً في قياس قدرة أوسمة شخصية غير متأثرة بالصدفة. ولكن إذا كثرت حالات الاختلاف بين الدرجتين في مجموعة من الأفراد، فإننا نكون أكثر ريبة في أن درجات جميع المفحوصين تعتمد إلى درجة كبيرة على عوامل الصدفة.

وكما رأينا في الفصل الثاني، فإن الطريقة المعتمدة للتعبير عن التشابه (أو الاختلاف) بين مجموعتين من الدرجات هو معامل الارتباط. ولأن عوامل الصدفة دائمًا موجودة إلى حدٍ ما فلن يستطيع الإخصائي الذي يبني الاختبار الحصول على معامل ارتباط تام بين مجموعتي الدرجات، أي لن يصل معامل الارتباط أبداً إلى (١,٠٠) الواحد الصحيح. ولكن إذا بُنِيَ الاختبار بطريقة جيدة وطبق بعناية تحت الظروف المناسبة، فإنه لن يكون من المستبعد أن نحصل على معامل ثبات يصل إلى حوالي ٠,٩٠ وفي بعض الأحيان ولأغراض خاصة، يمكن استخدام اختبار معامل ثباته أقل من ذلك بكثير. والمهم هو أن الإخصائي الذي سوف يقوم بعملية الاستدلال عن الأفراد أو الجماعات على أساس نتائج الاختبار لا بد أن يعرف إلى أي مدى هذا الاختبار ثابت حتى يستطيع أن يأخذ في اعتباره القيام بالتصحيحات اللازمة لنواحي القصور الموجودة في دقة الاختبار.

ومن المؤسف أن وصف ثبات الاختبار عن طريق ذكر معامل الارتباط بين صورتين متكافئتين للاختبار له مساوىء عديدة. وربما كان أهمها أن معامل

الارتباط، في حالة ما إذا كان مدى توزيع درجات المجموعة التي يحسب الثبات على أساسها كبيراً، يكون أعلى منه مما لو كان هذا المدى ضيقاً. ولذلك، فإنه من الضروري لمن يستخدم الاختبار، في حالة تقدير ثباته لغرض ما، أن يتأكد من أن معامل الثبات قد حسب في مجموعة لها من صفات التباين ما للمجموعة التي سيطبق عليها الاختبار. فعلى سبيل المثال، إن «اختبار التصنيف العام في الجيش»[١] على درجة كافية من الثبات لأن يميّز بين مستويات القدرة العقلية العامة للمجندين الذين يمثلون تنوعاً كبيراً في المواهب الطبيعية والتحصيل التعليمي. ولكن هذا الاختبار ليس ثابتاً بدرجة كافية لأن يميز في شيء من الوضوح بين طالب جامعي وطالب جامعي آخر: فالإخصائي النفسي الذي قد يستخدم هذا الاختبار في برنامج للإرشاد الجامعي سوف تكون أحكامه غير دقيقة إذا اعتمد على معامل ثباته الذي حسب في مجموعة من المجندين يتكون أفرادها من مختلف النوعيات. إنه يحتاج إلى معامل ثبات حسب في مجموعة مشابهة لطلبة الجامعة.

وهناك، لحسن الحظ، طريقة أخرى يمكن بها التحقق من مدى دقة أو عدم دقة درجات الاختبار، وهي طريقة حساب **الخطأ المعياري للقياس**. وقد سبق أن شرحنا المعنى العام **للخطأ المعياري** في الفصل الثاني. وهذه الأداة الإحصائية تعطي لنا تقديراً لمدى الاختلاف في درجات الفرد إذا طبق عليه الاختبار عدداً لا متناهياً من المرات ـ أو بمعنى آخر، إذا استطعنا أن نحصل على عدد كبير من «العينات» من درجات هذا الاختبار بأن نأخذها من «المجتمع» الكلي لدرجات هذا الاختبار. هذا المدى هو منطقة عدم الدقة الواقعة على كلا جانبي الدرجة. أما عن الأسس المنطقية والإحصائية لهذه الطريقة فهي معقدة جداً بحيث لا يمكن التعرض لها هنا، ولكن التفسير النهائي لها بسيط نسبياً بحيث يمكن فهمه.

إذا كان الخطأ المعياري لاختبار في الذكاء، على سبيل المثال، هو ٤,٦، فإن هناك احتمالاً بمقدار $\frac{٢}{٣}$ بأن الدرجة الحقيقية للفرد تقع في مدى ٤,٦ نقطة من الدرجة التي حصل عليها فعلاً. فإذا كان الفرد أ حصل على درجة ١٥٤، فإنه يمكن القول بدرجة من الثقة أن «درجته الحقيقية» ـ خالصة من تأثير

(١) The Army General Classification Test.

عوامل الصدفة - تقع بين ١٤٩,٤ و ١٥٨,٦ . وهذا يبنى على حقيقة أنه في أي توزيع اعتدالي نجد أن $\frac{٢}{٣}$ الحالات تقع في مدى انحراف معياري واحد من المتوسط. وكما أوضحنا سابقاً، فإن الخطأ المعياري هو نوع خاص من الانحراف المعياري. ولكن مازال هناك احتمال بمقدار $\frac{١}{٣}$ بأن «درجته الحقيقية» قد تكون أعلى أو أقل من ذلك. ونستطيع إذا شئنا، باستخدام التفكير الذي يعتمد على مفاهيم الخطأ المعياري، أن نقول أن هناك احتمالاً مقداره ٥٪ بأن تتراوح درجة الفرد بين ١٤٥ - ١٦٣. (ذلك لأنه في التوزيع الاعتدالي يقع ٩٥٪ من الحالات في مدى انحرافين معياريين من المتوسط). ولكننا عادة لا نقوم بمثل هذه التحديدات الدقيقة. إننا نستخدم الخطأ المعياري كدليل على مقدار عدم الدقة في تفسير الدرجات واتخاذ القرارات بالنسبة للأفراد. ففي مثالنا السابق، إذا كنا سوف نستخدم اختبار الذكاء هذا في اختيار الطلاب لدراسات متفوقة، وكان القرار أن كل طالب حصل على درجة ١٥٥ أو أعلى له الحق في هذه الدراسة، فإن الفرد أ، وقد حصل على ١٥٤، يبدو للوهلة الأولى كأنه غير مؤهل لها. ولكن المعلم الذي يأخذ في اعتباره المقدار العام لعدم الدقة في درجات الاختبار، سوف يرى أن الفرد أ قد يكون من المجموعة المؤهلة لهذه الدراسة، إذ أن من المحتمل جداً أن تكون درجة أ على الأقل ١٥٥. وفي هذه الحالة سوف يبحث المعلم في كثير من الدقة عن أدلة أخرى كثيرة للتحقق من حقيقة حالة الطالب أ قبل أن يقرر ما إذا كان يوضع في هذا الفصل الدراسي الخاص أم لا. وعلى ذلك، فإن فهم المعنى الحقيقي لثبات الاختبار سوف يكون له أثر واضح في استخدام الاختبارات في عملية اتخاذ القرارات. (وعندما نؤكد أن الاختبارات ليست ثابتة ثباتاً تاماً، فيجب أن نتذكر دائماً أن الطرق الأخرى في تقييم الناس ليست ثابتة كذلك. إن مؤلف الاختبار يخبرنا عادة عن مقدار عدم الدقة في اختباره. ولكن ليست لدينا، عادة، أية طريقة لمعرفة مقدار عدم الدقة في «أحكام» المعلمين).

إن الشخص الذي لا يفهم معنى «ثبات» الاختبار قد يتأثر تقييمه العام للاختبار بدرجة كبيرة بما يقال عموماً عن المعنى الدارج لكلمة الثبات. فكلمة «ثابت أو موثوق به»، كما تستخدم في كلامنا الدارج، تحمل معاني حسنة - فيقال رجل موثوق به (بمعنى «طيب»)، أو مؤسسة موثوق بها (بمعنى «مستقيمة»)، وإنتاج موثوق به (بمعنى «قيّم»). فمن المحتمل أن المعلمين، والعاملين في شؤون

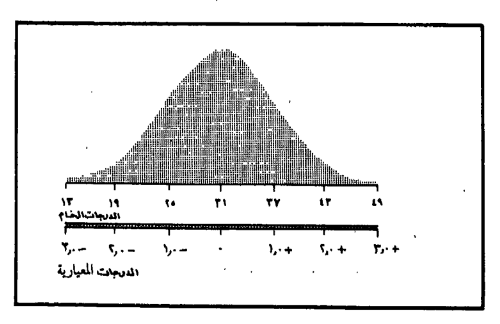
الأفراد، والإخصائيين الإكلينيكيين قد يستنتجون أن الاختبار الثابت هو، في الواقع، اختبار جيد صالح لأي غرض يفكرون فيه. وهذا خطأ يقود إلى أخطاء أكثر خطورة في الحكم على الناس. فإذا كانت هناك أدلة على أن اختباراً ما دقيق في قياسه لشيء ما، فإن هذا لا يعني كثيراً. أننا نحتاج إلى معلومات أكثر عن الاختبار. وقبل أن نعرف ماذا يقيس هذا الاختبار، لا نكون محقين في استنتاج نتائج عن الناس بناء على درجاتهم في الاختبار. وكما أوضحنا في الجزء السابق، فإن عملية البحث عن ماذا يقيس الاختبار بالضبط، هي عملية طويلة وشاقة. فكلمة «ثابت» لا تعني «صالح» لكل شيء. وعلى ذلك، فإن صدق الاختبار يظل أهم من ثباته.

المعايير والدرجات المشتقة

عند تقويم الاختبارات وتفسير الدرجات، يجب أن نأخذ في اعتبارنا، بالإضافة إلى صدق الاختبار وثباته، الوحدات المختلفة التي صيغت بها هذه الدرجات. فالدرجة التي تدل على عدد الأسئلة التي أجاب عنها الفرد إجابة صحيحة لا تقول لنا شيئاً كثيراً عن الفرد ما لم يكن عندنا معيار خاص للمقارنة. فإذا حصل تلميذ ما على درجة ٤٧ في الحساب، مثلاً، فإنه من المحتمل أن يكون أول سؤال يسأله والد التلميذ هو: «ما هي الدرجات الأخرى التي حصل عليها بقية التلاميذ؟» إن الدرجات المشتقة أو المحوّلة التي تستخدم في الاختبارات المقنّنة، سواء في الاستعدادات أو التحصيل أو الشخصية، إنما هي درجات صممت خصيصاً لمقارنة درجات الفرد بمعايير المجموعة. ويمكن أن نقوم بعملية اشتقاق الدرجات، أو تحويلها، بطرق عديدة.

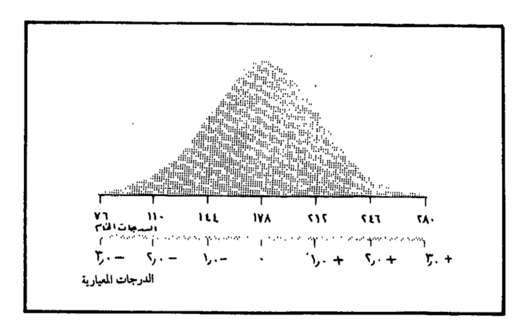
إن أبسط هذه الطرق هي أن نجهز معايير مئينية للمجموعة، ثم نحول درجة كل فرد إلى الرتبة المئينية المقابلة ـ ولكي نقوم بتجهيز جدول معايير مئينية لمجموعة ما، نقوم بقسمة عدد الأفراد الذين حصلوا على درجات أقل من كل درجة خام معينة، بالإضافة إلى نصف عدد الذين حصلوا على نفس هذه الدرجة، على العدد الكلي للمجموعة.. وفي مثالنا السابق عن اختبار الذكاء الذي حصل فيه التلميذ أ على درجة ١٥٤، فإذا وجدنا أن ٩١٪ من المجموعة حصلوا على درجة أقل من ١٥٤، فإن الرتبة المئينية للتلميذ أ في هذه المجموعة هي ٩١.

إن الطريقة الأخرى الشائعة لتحويل الدرجات الخام إلى درجات متكافئة تقوم بتوضيح مركز الفرد بالنسبة للمجموعة، تعتمد على استخدام المتوسط، والانحراف المعياري كأساس للمعايير. ففي التوزيع الاعتدالي هناك علاقة ثابتة بين البعد عن المتوسط والمساحة المحصورة تحت المنحنى. ويبين الشكل رقم (٦) ما هي هذه العلاقة. وبغض النظر عن كون الانحراف المعياري لأي مجموعة خاصة من الدرجات كبيراً أو صغيراً، فإنه يمكن استخدامه كوحدة ذات معنى لتحديد المسافات على قاعدة المنحنى. ففي التوزيع الاعتدالي، إذا ابتعدنا وحدة انحراف معياري واحد عن المتوسط، فإننا نجد أنفسنا مباشرة تحت النقطة التي يتغير فيها شكل المنحنى من التحدب إلى التقعُّر.. وعندما نبتعد عن المتوسط بمقدار ثلاث وحدات انحراف معياري نصل تقريباً إلى نقطة قريبة جداً من نهاية التوزيع. وعلى ذلك، فإنه يمكن، إذا شئنا، أن نعبر بوحدات الانحراف المعياري عن مقدار بعد كل درجة عن المتوسط، وبالتالي نستطيع أن نحدد بهذه الطريقة مكان كل فرد من التوزيع العام. فبدلاً من أن نذكر أن الدرجة ٢٥ قد حصل عليها فرد ما في المجموعة المبينة في الشكل رقم ٦، فإننا نستطيع أن نذكر درجته على أنها ـ ١,٠ انحراف معياري (أي تقل عن المتوسط بمقدار وحدة انحراف معياري واحدة). وبمجرد رؤية ذلك، فإن أي شخص له معرفة بمنحنى التوزيع الاعتدالي يمكنه أن يعرف بالضبط كم تقل درجة هذا الفرد عن المتوسط.



شكل ٦ ـ الدرجات المعيارية المقابلة للدرجات الخام في مستويات مختلفة لتوزيع متوسطه ٣١ وانحرافه المعياري ٦.

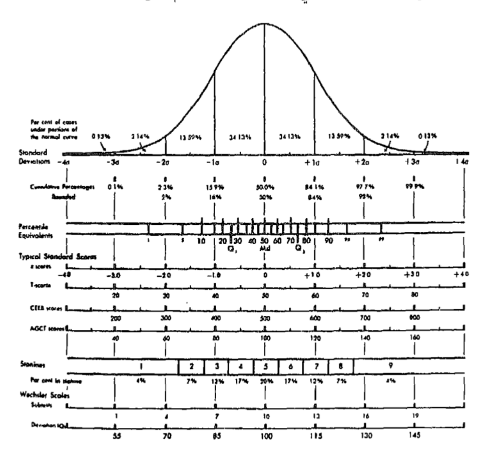
إن هذا النوع من تحويل الدرجات مكن علماء النفس من القيام بمقارنات لم يكن من الممكن القيام بها باستخدام الدرجات الخام. ففي توزيع اعتدالي آخر لدرجات اختبار ما (شكل ٧)، حيث المتوسط ١٧٨ والانحراف المعياري ٣٤، نجد أن الأرقام الموجودة على قاعدة المنحنى ممثلة للدرجات الخام مختلفة عن الأرقام المناظرة لها في شكل رقم (٦)، ولكن الدرجات المعيارية متماثلة. فالدرجة ١٤٤ في هذا الاختبار الثاني (١٧٨ - ٣٤) هي درجة تماثل في ضعفها الدرجة ٢٥ في الاختبار الأول، حيث أن كلاهما يساوي - ١,٠ انحراف معياري.



شكل ٧ - الدرجات المعيارية المقابلة للدرجات الخام في مستويات مختلفة لتوزيع متوسطه ١٧٨ وانحرافه المعياري ٣٤.

ورغبة في التبسيط. فإننا كثيراً ما نعبّر عن الدرجات المعيارية الأساسية بطرق أخرى كثيرة. فمثلاً، من أجل الحصول على درجات مقسمة إلى وحدات أصغر، يمكن أن نضرب كل الأعداد × ١٠ بحيث يتراوح مدى الدرجات من - ٣٠ إلى + ٣٠ بدلاً من - ٣ إلى + ٣. ثم من أجل إزالة الدرجات السالبة، يمكن إضافة ٥٠ بحيث تتراوح الدرجات من ٢٠ إلى ٨٠، مما يجعل المتوسط ٥٠ بدلاً من الصفر. ومثل هذه التعديلات لا تغير أبداً في العلاقات الأساسية بين الدرجات. وفي الشكل رقم (٨) نرى كيف ترتبط مجموعة من هذه الدرجات المشتقة وكذلك التدريج المئيني

بالتوزيع الاعتدالي الأساسي. فالقيم على الخطوط الرأسية في الشكل والموجودة على مختلف الخطوط الأفقية تحت المنحنى الاعتدالي سوف تكون متساوية إذا كانت المجموعات المعيارية التي وضعت هذه القيم على أساسها متشابهة.



شكل ٨ ـ العلاقات بين مختلف أنواع الدرجات المشتقة.

(From H. G. Seashore, **Methods of Expressing Test Scores.** Psychological Corporation Test Service Bulletin, No. 48 New York, 1955.)

وبالإضافة إلى الرتب المئينية والدرجات المعيارية، هناك طرق أخرى عديدة لجعل الدرجات مفهومة: مثل معايير العمر الزمني والفرقة الدراسية في اختبارات التحصيل المدرسية وهي لا تحتاج إلى شرح خاص، أو نسبة الذكاء (I Q) التي سوف نتناولها في الفصل التالي.

عدم التحيز (الموضوعية)

لقد ثار جدل كثير، كما أجريت بحوث كثيرة، طوال السنوات الماضية، حول موضوعية الاختبارات العقلية بالنسبة للمفحوصين الذين هم من أجناس

وجنسيات مختلفة، ومن مستويات اجتماعية متباينة. وفي منتصف الستينات من هذا القرن، عندما تركزت الأضواء على الأحوال السيئة للفقراء في الولايات المتحدة الأمريكية، أصبحت هذه القضية ذات أهمية كبيرة. وبالتالي فقد أصبح كل من يستخدم الاختبارات في مهنتهم ـ كالمعلمين والإداريين في المدارس، ومديري الأعمال، ورؤساء الوكالات الحكومية ـ أصبح كل هؤلاء يهتمون بموضوعية الاختبار المستخدم بالنسبة للمفحوصين، وذلك بالإضافة إلى الصدق والثبات والتقنين المناسب.

والموضوعية مفهوم معقد، مثل غيره من المفاهيم التي شرحناها في هذا الفصل. وذلك لأن كل سمات الإنسان قد تكونت من خلال الخبرة، وبالتالي فإن الأشخاص ذوي الخلفية المتباينة سوف تختلف استجاباتهم حتماً للأسئلة التي يتضمنها أي اختبار ما. وليس هناك ما يسمى بالاختبار الخالي من العوامل الحضارية. إن ما يجب أن يفعله مؤلف الاختبار أو من يستخدمه، إذن، هو أن يطبق الاختبار على مجموعات مختلفة من الأفراد الذين سيطبق عليهم هذا الاختبار في نهاية الأمر، لكي يرى ما إذا كانت بعض الجماعات تحصل على درجات أقل من غيرها، بل وكذلك لكي يرى كيف ترتبط درجاتها بالمحكات ذات العلاقة.

فعلى سبيل المثال، لو أن معظم أفراد جماعة من جماعات الأقلية يميلون إلى الحصول على درجات منخفضة في اختبار للاستعدادات صمم للتنبؤ بمدى نجاح المتدربين في برنامج تدريبي لمهنة ما، ولو أن نفس الأفراد يميلون إلى الحصول على درجات منخفضة على مقياس المحك (بحيث يكون معامل الارتباط بين درجات الاختبار ودرجات مقياس المحك ٠,٥٩ مثلاً)، فإن الاختبار يمكن اعتباره موضوعياً حتى ولو كانت نتائجه ليست من صالح هذه الجماعة. وما يجب أن ينتبه إليه دائماً من يستخدم الاختبار في أغراض اختيار الأفراد، هو عدم تشابه الصدق التنبئي لهذا الاختبار في مجموعة الأقلية مع الصدق التنبئي في المجموعات الأكثر تمثيلاً لطالبي العمل الذين أجريت عليهم من قبل دراسات الصدق. فلو كان معامل الارتباط بين درجات الاختبار ودرجات المحك، في مثالنا السابق، ليس ٠,٥٩ بل ٠,٠٦، فإننا سوف نستنتج أن هذا الاختبار ليس موضوعياً بالنسبة لهذه الجماعة من طالبي العمل، بغض النظر عن مدى صدقه في جماعات أخرى. والطريقة الوحيدة للتأكد من أن الاختبار الذي سوف

نستخدمه لتقدير قدرة ما عند جماعة من طالبي العمل الجدد هو اختبار موضوعي بالنسبة لأفراد هذه الجماعة، هي أن نجري مجموعة من الدراسات لبحث العلاقة بين درجات الاختبار والأداء في العمل.

وعندما تكون معاملات صدق الاختبار عالية في جماعة ما «مهضومة الحق (أو محرومة)»، فإن الاختبار يمكن أن يكون أداة مفيدة تساعد على التصدي للتمييز والتفرقة، ويفتح الفرص أمام الأفراد. وتفيد الاختبارات في كثير من الأحيان في اكتشاف المواهب في الحالات التي قد لا يلتفت إليها لو لم تستخدم الاختبارات، وذلك بسبب التعصب الشعوري أو اللاشعوري لدى القائمين بالمقابلة ضد طالبي العمل من السود أو السمر أو من ذوي الملابس الرثة. ونعود ونقول أنه عندما نتحدث عن موضوعية الاختبار يجب أن نتذكر أن الوسائل الأخرى للحكم على قدرات الأفراد وإمكاناتهم هي أيضاً عرضة لكي تكون غير موضوعية مثل الاختبارات ـ وربما تكون أقل موضوعية من الاختبارات.

ومن المهم أيضاً في بعض الأحيان أن نعرف علاقة أي اختبار جديد بالاختبارات الأخرى في مجموعة مختلفة من المفحوصين، بحيث يمكن، عن طريق عملية التحليل العاملي، توضيح طبيعة بناء القدرة في هذه الجماعة. وباختصار، فإن الحكم على موضوعية الاختبار يجب أن يبنى على نتائج كل جماعة على حدة، وكذلك بالنسبة لكل غرض من أغراض القياس. وسوف نعود مرة أخرى لهذا المفهوم في فصول تالية عندما نناقش أنواعاً معينة من الاختبارات.

تصنيف الاختبارات

إذا أردنا أن نأخذ في الاعتبار أسس وأساليب الاختبارات النفسية في شيء من التفصيل، فلا بد إذن من أن نصنف الاختبارات في فئات رئيسية قليلة. وليست هناك طريقة مرضية تماماً للقيام بذلك، ولكن التصنيف الأكثر قبولاً هو التصنيف الذي يعتمد على ما صمم الاختبار لقياسه. فمن الممكن أن نميز بصفة أساسية بين اختبارات القدرات التي نحدد بوساطتها ماذا يمكن أن يقوم به الفرد، وبين اختبارات الشخصية التي نحدد بوساطتها ماذا يشعر به الفرد، أو ماذا يريد، أو ماذا يقلقه. أما عن اختبارات القدرات فإنه من المفيد أن نميز بين اختبارات الذكاء والاختبارات الخاصة ببعض القدرات المختلفة المحدودة.

لقد ناقشنا في هذا الفصل عدة مفاهيم وطرق تنطبق على كل أنواع

ـــختبارات. وهناك مفاهيم وطرق أخرى تنطبق على نوع واحد من الأنواع الثلاثة من الاختبارات - الذكاء، أو القدرات الخاصة، أو الشخصية - أكثر من انطباقها على النوعين الآخرين. وسوف نتناول هذه المفاهيم والطرق في الفصول التالية.

الفصل الرابع

اختبارات الذكاء

إن محاولة قياس الذكاء الإنساني قد أخذت من الجهد الطويل المكثف والمستمر أكثر من أي موضوع آخر من موضوعات القياس النفسي. ففي دراسة الذكاء يتفاعل الفضول الفلسفي مع المطالب العملية. ولمدة قرون والإنسان يفكر في الفروق الهائلة في القدرة العقلية التي تميز إنساناً من طراز سقراط عن الفرد العادي، أو تميز الطفل الأبله عن الطفل العادي. وقد أثير السؤال عن طبيعة هذه الفروق هل هي فطرية أم مكتسبة - هل العوامل البيولوجية أم العوامل التربوية أكثر تأثيراً في إظهار هذه الفروق؟ وأخيراً، بسبب الحاجة العملية في المجتمعات المتمدينة، أصبح الأمر أكثر إلحاحاً لإيجاد طريقة ما لتقدير ذكاء الأفراد تقديراً دقيقاً على قدر الإمكان. فظهور التعليم الشامل أعطى الفرصة لجميع الأطفال، موهوبين وغير موهوبين، في الذهاب إلى المدرسة. وبعض هؤلاء التلاميذ لم يتمكنوا من إجادة البرامج التي أعدها لهم المتخصصون في التربية، في حين أن البعض الآخر تمكن من هذه البرامج بسرعة فائقة؛ بل إن هذا البعض اهتم بالتجارب العلمية أو التأملات الفلسفية في وقت مبكر، وقبل

أن يصلوا إلى تمام نضجهم البدني. وبالتالي، فقد أصبح من الواضح احتياج المعلمين لأن يميزوا بين الإمكانات العقلية المختلفة حتى يمكنهم تعليم الأطفال في صورة مناسبة. وقد ظهرت نفس المشكلات في التنظيمات الحربية والصناعية فيما يتعلق بوضع الأفراد في أعمال ملائمة.

وعندما بدأ الإخصائيون النفسيون جهودهم لقياس الذكاء، وذلك قبل بداية القرن العشرين بقليل، كانت تنقصهم المعرفة الدقيقة بطبيعة هذه القدرة. وكما رأينا عند مناقشة الصدق في الفصل السابق، فإن مثل هذا الموقف ليس أمراً غير عادي، كما أنه ليس بالضرورة أمراً سيئاً، إذ أن مجرد فكرة عامة، حتى وإن كانت غير واضحة، عن سمة إنسانية قد تساعد في إرشاد من يقوم ببناء الاختبار عندما يختار بنود الاختبار ويخطط لدراسات الصدق. ثم بتراكم المعلومات حول ما ترتبط به درجات الاختبار، يصبح قادراً على وضع تعريف أكثر وضوحاً لما يقيسه الاختبار، ومن ثم يعدل الاختبار ويحسنه.

إن أول اختبار ذكاء للأطفال، وهو مقياس بينيه وسيمون Binet- Simon scale، قد جاء بعد فترة طويلة من الملاحظة الدقيقة التي قام بها بينيه للفروق بين عمليات التفكير عند الأطفال. فقد أجرى دراسات على بناته الصغيرات، وناقش المعلمين في كيفية حكمهم على الإمكانية العقلية، كما قام بالعديد من التجارب قبل أن يقدم اختبار الذكاء للناس. فهذا العمل قد أمدّ بينيه بمعلومات معينة عن ماذا يجب أن يكون عليه اختبار الذكاء. ففي المقام الأول، رأى بينيه أن الاختبار يجب أن يعكس النمو السريع في الإمكانية العقلية الذي يحدث في فترة الطفولة. وثانياً - لأسباب عملية إن لم يكن لسبب آخر - يجب أن يقيس الخصائص التي يعتمد عليها المعلمون في التفرقة بين الأطفال «النابهين» و «الأغبياء» داخل الفصول الدراسية. وقد كان تعريف بينيه المبدئي للذكاء كما يقيسه الاختبار الأول له هو «الميل إلى اتخاذ وجهة محددة والمحافظة عليها؛ والقدرة على التكيف من أجل الوصول إلى الهدف المطلوب؛ والقدرة على النقد الذاتي». وكما سبق إن ذكرنا، فإن نشر اختبار الذكاء في ذلك الوقت بالذات قد تم بناء على طلب من المسؤولين التربويين في باريس لإيجاد طريقة تساعد على التمييز بين الأطفال الأغبياء بالفعل والأطفال الذين هم كسالى فقط.

وتطور الأمر بعد ذلك بسرعة، فقد أعيد النظر في اختبار بينيه وسيمون

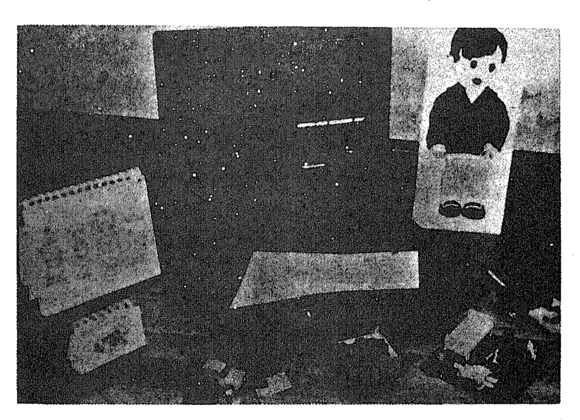
مرتين، مرة في عام ١٩٠٨، والأخرى في عام ١٩١١، كما ترجم وعدل في الكثير من البلاد. ففي الولايات المتحدة قام لويس م. ترمان Lewis M. Terman من جامعة ستانفورد بإعداد الصورة الأمريكية من الاختبار، وكانت النتيجة اختبار ستانفورد ـ بينيه Stanford- Binet في عام ١٩١٦، ثم عدل في عام ١٩٣٧، وأعيد النظر فيه مرة ثالثة في عام ١٩٦٠. ولمدة نصف قرن من الزمان ظل اختبار ستانفورد ـ بينيه معياراً معترفاً به لقياس الذكاء مثل مقاييس الأطوال والأوزان. أما في الوقت الحاضر، فقد ظهرت بدائل عديدة من الاختبارات أكثر مما كان موجوداً في السنوات السابقة، وقل الميل إلى الرجوع إلى اختبار ستانفورد ـ بينيه في جميع حالات قياس الذكاء. ولكن الدور الذي أداه هذا الاختبار في تطوير المفاهيم النظرية والتطبيقات العملية للاختبارات العقلية فهو أمر لا يمكن إهماله.

الخصائص المميزة لاختبارات بينيه

على الرغم من أن الصور المعدلة المتتالية من اختبار بينيه قد اختلفت عن بعضها البعض، كما اختلفت عن اختبار بينيه ـ سيمون الأصلي، إلا أن هناك مجموعة من الملامح مشتركة بين جميع هذه الصور المعدلة. فأولاً، إنها جميعاً مقاييس. وهذا يعني أن الأسئلة والمهام التي تحتوي عليها قد صنفت على أساس درجة صعوبتها. فالمختبر يبدأ بتوجيه الأسئلة السهلة، ثم ينتقل بعد ذلك إلى الأسئلة الأصعب فالأصعب مع تقدم الاختبار. وتعتمد درجة الطفل أساساً على مقدار أعلى مستوى يستطيع الطفل أن يصل إليه في هذا السلم المتدرج أكثر مما تعتمد على سرعته وطلاقته (على الرغم من أن السرعة والطلاقة مطلوبتان في بعض الاختبارات الفردية). وابتداء من تعديل اختبار بينيه الذي حدث في عام ١٩٠٨، جمعت الاختبارات بناء على مستويات العمر. وقد بذلت جهود كبيرة لمعرفة معايير العمر الزمني، حيث أنه من أجل تعيين اختبار ما لعمر زمني محدد فإنه من الضروري أن نختبر كيفية تعامل الأطفال من أعمار مختلفة مع هذا الاختبار. فالاختبارات التي تعيّن لسن العاشرة، على سبيل المثال، يجب أن تحتوي على مهام يستطيع عدد قليل فقط من أطفال سن التاسعة القيام بها، ولكن يستطيع القيام بها أغلبية أطفال سن العاشرة، وعدد أكبر من أطفال

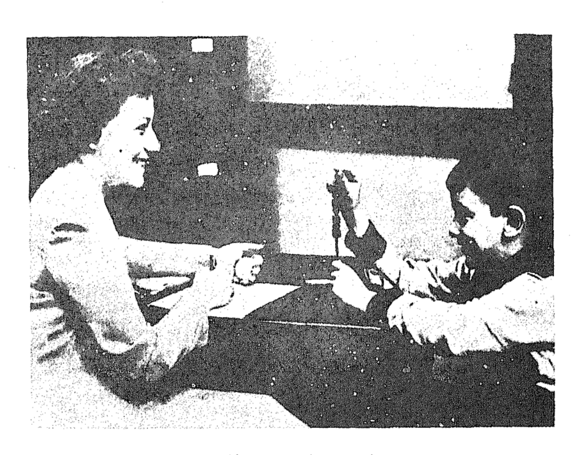
الحادية عشرة. ولذلك، فإن اختباراً للذكاء من نوع اختبار بينيه يعتبر مصمماً لقياس النمو العقلي.

والخاصة الثانية المشتركة بين اختبارات بينيه هي أننا نحصل عن طريقها على قياس عام شامل للذكاء بدلاً من تحليل لقدرات خاصة متفرقة. فمن الواضح من تعريف بينيه أنه كان يرى أن الذكاء إمكانية مركبة قادرة على معالجة أنواع عديدة من الأعمال، أكثر من كونه مجموعة من القدرات المختلفة للقيام بأعمال معينة. وتمشياً مع هذا الرأي، فإن اختبارات مستوى عمري معين قد تتطلب معرفة باللغة (الكلمات أو المفردات)، وإدراك التفاصيل في الصور، والتفكير، والتذكر المباشر. ولم تكن هناك أية محاولات للفصل بين الأنواع المختلفة من القدرات. ويبين الشكل رقم ٩ الأدوات المختلفة المستخدمة في اختبار ستانفورد ـ بينيه لعام ١٩٦٠، وهو يعطينا فكرة عن تنوع هذه الأدوات.



شكل ٩ ـ الأدوات المستخدمة في اختبار ستانفورد ـ بينيه.

والخاصة الثالثة التي تميز الاختبارات التي من نمط اختبارات بينيه عن الاختبارات الأخرى هي أنها صممت لتطبق بصورة فردية بواسطة إخصائيين على درجة عالية من المهارة (أنظر الشكل رقم ١٠) . إذ أن الأمر يحتاج إلى



شكل ١٠ ـ تطبيق اختبار ستانفورد ـ بينيه

(From the City College of New York Educational Clinic)

مهارة ـ أكثر مما قد يُعتقد ـ في تكوين علاقة طيبة واتجاه مشجع مع الطفل المفحوص في نفس الوقت الذي تعطى له التعليمات، علمًا بأن هذه التعليمات قاسية. فعلى سبيل المثال، على الفاحص ألا يغير أبداً في نص كلمات الأسئلة مهما كان ـ فأي تعديل ولو بسيط من وجهة نظره قد يؤدي إلى جعل السؤال أصعب أو أسهل مما كان عليه الأمر في المجموعة التي اشتقت منها معايير العمر الزمني. وقد يقوم الفاحص بالثناء على المفحوص لكي يحثه على بذل أقصى مجهوده، ولكن لا يجب أن يفعل ذلك أكثر من اللازم أو في أوقات غير مناسبة. وقد يقوم الفاحص بتكرار بعض الأسئلة، ولكنه لا يقوم بتكرار البعض الآخر. وإذا لم يكن من الواضح أن إجابة ما تطابق المواصفات الموضوعة لمستوى عمري معين، فقد يقوم الفاحص بتوجيه سؤال ما، ولكن يجب أن يحرص على ألا يعطى المفحوص أية إشارة إلى الإجابة الصحيحة. إن القدرة على التصرف بطريقة طبيعية وودية وفي هدوء، أثناء القيام بإجراءات الاختبار وتوجيه الأسئلة المحددة بطريقة جامدة، إنما تتطلب قدراً كبيراً من التدريب. ومن الواضح أنه

ليس في استطاعة أي شخص توفرت لديه الأدوات المستخدمة أن يقوم بإجراء اختبار بينيه.

وأخيراً، فإن النظام المتبع في إعطاء الدرجات في جميع اختبارات بينيه إنما يرتبط بمعايير العمر. فالعمر العقلي لطفل ما يدل على المجموعة العمرية التي يتناسب أداؤه معها تماماً. فعلى سبيل المثال، إن العمر العقلي الذي يساوي ٩ سنوات وستة شهور يعني أن النمو العقلي العام لهذا الطفل يشبه النمو العقلي العام عند الطفل العادي الذي عمره تسع سنوات ونصف. وعلى ذلك، فإن العمر العقلي يشبه إلى حدٍ ما الأرقام التي تدل على حجم ملابس الأولاد والبنات، إذ أنه يمكن معرفة حجم جسم طفلة ما عندما نقول أنها تلبس رقم ١٠. وبالمثل يمكن أن نعرف «حجم» قدرتها العقلية عندما نقول أن عمرها العقلي ١٠.

معنى نسبة الذكاء.

إن اختبار ستانفورد ـ بينيه والاختبارات الأخرى التي على نمط اختبار بينيه تمكننا أيضاً من حساب **نسبة الذكاء** (I Q). وحتى التعديل الذي حدث في عام ١٩٦٠، عندما استخدمت الوسائل الإحصائية الدقيقة، كان الحصول على نسبة الذكاء يتم بمجرد قسمة العمر العقلي للطفل على عمره الزمني ثم ضرب الناتج × ١٠٠ ليتم التخلص من الكسور. فعلى سبيل المثال، إذا كان العمر العقلي لطفلٍ ما هو ١٠، وعمره الزمني ٩، فإن نسبة ذكائه سوف تساوي ١١١ ($\frac{١٠}{٩}$ × ١٠٠). إن نسبة الذكاء توضح مدى مناسبة نمو الطفل بالنسبة لعمره (كم هو كبير أو صغير بالنسبة لعمره). إنه يخبرنا عن «معدل» النمو العقلي له حتى لحظة إجراء الاختبار عليه.

ومن المؤسف أن مصطلح «نسبة الذكاء» أصبح كثير الشيوع، وقد استخدم في أغراض كثيرة لا مبرر لها. وما زالت الأخطاء في فهم معناه، في الواقع، شائعة حتى الآن. فنسبة الذكاء ليست هي المؤشر الأهم الوحيد على الإمكانية العقلية، كما يفترض في كثير من الأحيان. كما أنه ليس مقياساً لكمية من أي شيء، ولكنه بكل بساطة طريقة لمعرفة متوسط معدل النمو العقلي للطفل. وفي مرحلة من مراحل البحوث والدراسات في الذكاء كان يعتقد أن معدلات النمو سوف تكون ثابتة خلال مرحلة الطفولة إلى درجة كافية بحيث يمكن اتخاذها

أساساً للتنبؤ الدقيق بدرجة الذكاء في مرحلة الرشد. ولكن حينما تجمعت لدينا معلومات وفيرة، أصبح واضحاً أن «ثبات نسبة الذكاء» أمر بعيد تماماً عن كونه قضية مطلقة. وذلك، من ناحية، لأنه حتى في حالة الاختبارات المقننة والتي وضعت بعناية ودقة للمستويات العمرية كما هو الحال في تعديل عام ١٩٣٧ لاختبار ستانفورد ـ بينيه، نجد أن نسب الذكاء في المستويات العمرية المختلفة غير قابلة تماماً للمقارنة الإحصائية. فالطفل الذكي جداً يمكن أن يحصل على نسبة ذكاء في سن الثانية عشرة أعلى من نسبة الذكاء التي حصل عليها وهو في سن السادسة، حتى ولو لم يتغير معدل نموه على الإطلاق، وذلك لسبب بسيط، وهو أن تشتت توزيع نسب الذكاء بين أطفال سن الثانية عشرة أكبر من تشتت توزيع نسب ذكاء الأطفال في سن السادسة. ويصبح الاختلاف في معنى نسب الذكاء العالية والمنخفضة من عمر إلى آخر، أكثر تعقيداً في حالة الاختبارات التي لم تقنن بالدقة الكافية.

وهناك عيب آخر وهو أن نسبة الذكاء لا تصلح لوصف ذكاء الراشدين. فالنمو العقلي لدى الراشدين، شأنه في ذلك شأن النمو الجسمي، لا يمكن التنبؤ بمعدله بطريقة منتظمة كما هو الشأن في حالة الأطفال. فابتداء من منتصف العقد الثاني تصبح المعايير العمرية عديمة المعنى نسبياً. فليس هناك معنى لأن نقول أن شخصاً ما عمره الزمني ٢٠ سنة وعمره العقلي ٢٥، ذلك لأنه ليس هناك فرق بين استجابات شخص عمره ٢٠ سنة واستجابات آخر عمره ٢٥ سنة لأسئلة اختبارات الذكاء. ولذلك فإن الدرجة المعيارية تحل، في الواقع، محل نسبة الذكاء في حالة الراشدين. فهذه الدرجة توضح أن الفرد البالغ يحتل في توزيع الراشدين نفس المكان الذي يحتله طفل له نفس نسبة الذكاء في توزيع مماثل للأطفال. وبدراسة الإخصائيين النفسيين للتعقيدات التي تتضمنها درجات نسب الذكاء، أصبحوا يدركون أن جميع نسب الذكاء للأطفال وللراشدين يمكن أن تفسر على اعتبار أنها درجات معيارية. إذ أن نسبة ذكاء فرد ما إنما تقول لنا كم عدد الانحرافات المعيارية التي يبتعد بها الفرد فوق المتوسط أو تحته.

ومن الأشياء المفيدة جداً التي يمكن أن يتعلمها الطالب هي ألا يعير اهتماماً كبيراً لنسبة الذكاء في صورة رقمية. إن كثيراً من الاختبارات المقننة تقنيناً جيداً تمدنا الآن بجداول معايير يمكن الاستعانة بها في الحصول على درجات أخرى مشتقة ـ وهي درجات توضح للفرد مباشرة مركزه في الجماعة التي يرغب

أن يقارن نفسه بها، سواء كانت هذه الجماعة، مثلاً، أطفالاً في سن السابعة، أو طلبة في المدارس الثانوية أو مجندين في الجيش. فإن اختبار ستانفورد ـ بينيه المعدل عام ١٩٦٠، على الرغم من أنه استمر في استخدام مصطلح نسبة الذكاء، إلا أنه لا يستخدم الطريقة القديمة في حسابها ـ أي قسمة العمر العقلي على العمر الزمني. ولكن بدلاً عن ذلك، فإن الفاحص يستخدم جداول تبين مباشرة مدى اختلاف درجة الطفل عن المتوسط الحسابي لدرجات مجموعة من الأطفال ممثلة للأطفال في مثل سنه وهذه الدرجة المشتقة تسمى نسبة ذكاء، ولكنها في الواقع درجة معيارية. إن مفهوم نسبة الذكاء قد أفادنا جيداً في الأيام الأولى من نشأة اختبارات الذكاء، أما الآن فإنه لا يستخدم كثيراً.

اختبارات وكسلر

بجانب لويس ترمان الذي صمم الصورة الأصلية لاختبار ستانفورد ـ بينيه، ومود ميريل Maud Merrill الذي ساعده في تعديل ١٩٣٧ وأعد تعديل ١٩٦٠، كان هناك أمريكي آخر ذو أهمية واضحة في قياس الذكاء، وهو دافيد وكسلر David Wechsler. ففي ١٩٣٩ نشر وكسلر مجموعة من الاختبارات الفردية المقننة صممت خصيصاً للراشدين، وسميت مقياس وكسلر ـ بلفيو (والجزء الثاني من الاسم إنما هو تكريم لمستشفى بلفيو في نيويورك حيث كان يعمل وكسلر). وسرعان ما استخدم هذا الاختبار بكثرة نتيجة تزايد الطلب على تقدير ذكاء الملايين من الراشدين مع النمو غير المتوقع في ميدان علم النفس الإكلينيكي أثناء الحرب العالمية الثانية وبعدها، حيث أن اختبار بينيه بتعديلاته المتعددة قد سدّ حاجات المؤسسات التي تهتم بالطفولة، ولكنه لم يكن مرضياً كلية بالنسبة للراشدين.

ومنذ الحرب العالمية الثانية تمكن وكسلر من بناء مقياس خاص بالأطفال على نفس نمط مقياس وكسلر ـ بلفيو الأصلي، كما أنه عدل أيضاً مقياس الراشدين بدقة، وبالتالي فإن التعديلين الشائعي الاستعمال الآن لاختبارات وكسلر هما مقياس وكسلر لذكاء الراشدين (WAIS) الذي نشر عام ١٩٥٥، ومقياس وكسلر لذكاء الأطفال (WISC) الذي نشر عام ١٩٤٩. وعلى الرغم من أن هناك بعض التداخل بين كلا المقياسين بدرجة ما، إلا أن الأول صمم أساساً لسن ١٦ سنة فما فوق، والثاني صمم لسن ١٥ سنة وما تحتها.

وقد تضمن مقياسا وكسلر نفس نوعية الأسئلة والمهام التي استخدمها بينيه وترمان والآخرون، إلا أنه ينظمها بطريقة مختلفة. فبدلاً من أن يجمعها حسب مستويات العمر، فقد جمعها حسب نوع السؤال أو المهمة، مرتباً البنود النوعية في كل مجموعة بناء على درجة الصعوبة. فعلى سبيل المثال، كل الأسئلة الحسابية تجمع في اختبار فرعي واحد، وكل المهام الخاصة ببناء المكعبات في اختبار فرعي آخر، وهكذا. وكذلك نجد أن الاختبارات الفرعية قد جمعت في صنفين رئيسيين هما اختبارات لغويةوأخرى عملية. والاختبارات اللغوية في اختبار الراشدين هي: المعلومات،والفهم،وتذكر الأرقام، والمتشابهات، والحساب، والحصيلة اللغوية. أما الاختبارات العملية فهي ترتيب الصور، وإكمال الصور وبناء المكعبات، وتجميع الأشياء، ورموز الأرقام. (والاختبارات الفرعية في الاختبار الخاص بالأطفال تشبه ذلك مع بعض التعديلات البسيطة القليلة)، فالاختبارات اللفظية تتطلب من المفحوص أن يشرح معاني الكلمات، وأن يعطي الإجابات الصحيحة لأسئلة المعلومات العامة مثل «ما هي عاصمة النرويج؟»، وأن يقوم بإجراء بعض العمليات الحسابية بدون استخدام القلم والورقة، ويوضح الشكل رقم ١١ شخصاً يؤدي اختبار بناء المكعبات وهو من اختبارات المقياس العملي.

وقد أعد وكسلر معايير خاصة بكل اختبار من الاختبارات الفرعية، وبذلك يمكن مقارنة الراشد أو الطفل في كل فئة من الأسئلة بمجموعة من الأفراد تمثل المجتمع الكلي. وفضلاً عن ذلك، فإنه يمكن أن نجمع الدرجات المعيارية على كل اختبار من الاختبارات الفرعية لنحصل على درجة كلية في المقياس اللفظي، ودرجة كلية في المقياس العملي، ودرجة كلية في المقياس ككل. وأخيراً، بمقارنة هذه الدرجات الثلاث الكلية بجداول المعايير الخاصة بعمر المفحوص يمكن أن نعرف نسبة ذكائه اللفظي، ونسبة ذكائه العملي، ونسبة ذكائه الكلي . ونسبة الذكاء هذه، مثل نسبة الذكاء التي نحصل عليها من تعديل ترمان وميريل عام ١٩٦٠، ليست نسبة حقيقية بين العمر العقلي والعمر الزمني، وإنما هي تشير إلى مدى علاقة الفرد بالفرد المتوسط في توزيع مجموعته العمرية باستخدام وحدات الانحراف المعياري. (انظر شكل ٨ ص ٦٣ لمعرفة العلاقة بين نسبة الذكاء في مقياس وكسلر ووحدات الانحراف المعياري الأساسية في كثير من أنظمة الدرجات المشتقة).



شكل ١١ ـ عينات من المواد المستخدمة في اختبار الذكاء بواسطة مقياس وكسلر لذكاء الراشدين.
(From The Psychological Corporation.)

وعندما أصبحت هذه المقاييس متوفرة لأول مرة، كان من المأمول أن تساعد الفروق بين درجات الاختبارات الفرعية في تشخيص وتقويم الوظائف العقلية مما من شأنه أن يعين علماء النفس الإكلينيكيين والمعلمين الذين يهتمون بمعرفة الكثير عن الفرد، الأمر الذي لا توفره الدرجة الكلية. وقد كان يظن أن تحليل مواطن القوة والضعف في الاستجابات للاختبارات المختلفة يمكن أن يساعد في الإجابة على بعض الأسئلة الخاصة مثل: كيف يصنف الفرد «أ» بين المرضى النفسيين؟ أو هل الفرد « ب » يعاني من تلف من نوع ما في المخ؟ أو ما هي العوائق التربوية التي تحول دون تقدم الفرد « جـ »؟ غير أنه بمضي الوقت أصبحت الدلالة التشخيصية لنماذج الدرجات أو بروفيلاتها أمراً يقابل بكثير من الحذر. فإننا نعرف الآن أنه توجد عدة أسباب محتملة تفسر لماذا يحصل الفرد

«ا» ، مثلاً، في اختبار الحساب على درجة أقل من درجاته في الاختبارات الأخرى. فقد يدل ذلك على القلق، غير أن ذلك قد يدل أيضاً على أن المفحوص لم يدرس مادة الحساب على الإطلاق. ولا يزال المختصون في علم النفس الإكلينيكي يحللون بروفيل درجات الفرد ويمحصون إجاباته على بعض الأسئلة الخاصة محاولين البحث عن دلالات تساعدهم على فهمه، ولكنهم قد تعلموا أن يعتبروا أن ما يحصلون عليه من أفكار بهذه الطريقة إنما هي مجرد فروض تحتاج أن تراجع بمضاهاتها بالمعلومات الأخرى المتوفرة عن الفرد.

والفروق بين درجات الاختبارات اللفظية والاختبارات العملية ربما تكون ذات معنى أكثر من الاختلافات بين درجات بعض الاختبارات الفرعية. ولكن لا بد أن نتوخى الحذر في تفسير هذه الفروق أيضاً. فليس أي من هذين المقياسين ثابتاً أو دقيقاً جداً بحيث أن الفروق البسيطة بينهما يمكن أن تؤخذ على محمل الجد. ولكن إذا حصل فرد ما على ١٥ نقطة أو أكثر على مقياس منهما أكثر من المقياس الآخر (١٥ نقطة هي الانحراف المعياري لتوزيع نسبة الذكاء في حالة اختبارات وكسلر)، فإنه قد يكون من المفيد محاولة تعليل هذا الفرق، وأن نأخذ في الاعتبار ما قد يعني ذلك عند التخطيط لمستقبل هذا الفرد - سواء كان هذا التخطيط يتعلق بالعلاج الإكلينيكي، أو التحصيل التربوي، أو الاختيار المهني. فمن الواضح أن المقياس اللفظي أكثر ارتباطاً بالحياة المدرسية من المقياس العملي. فالدرجة المتدنِّية في الاختبارات اللفظية التي تصاحب درجة عالية من الاختبارات العملية تعكس في كثير من الأحيان بعض التخلف التربوي. كما أن درجة الاختبارات اللفظية تساعد أيضاً على التنبؤ بدرجة أدق كثيراً من درجات الاختبارات العملية بمدى نجاح الفرد في المواقف المدرسية في المستقبل. وفضلاً عن هذا الفرق في التنبؤ بالنجاح المدرسي، فإن هناك العديد من العوامل الأخرى قد تكون مسؤولة عن الفرق من نصفي الاختبار (اللفظي والعملي). فعندما يقابل الفاحص مثل هذا الاختلاف فإنه يبحث في سجل المفحوص عن تعليل لذلك، سواء فيما يكتبه المفحوص عن نفسه أو فيما يكتبه غيره عنه، وهذا بالطبع سوف يساعد من يتعامل مع هذا المفحوص على أن يفهمه بصورة أدق. ففي التقرير الذي يكتبه الفاحص عن نتائج اختبار المفحوص، لا يكتفي الفاحص عادة بتسجيل الدرجات، ولكنه يضيف أيضاً بعض التحليلات الوصفية لما قد تعنيه هذه الدرجات.

وعلى العموم، فإن المعلومات الأكثر قيمة التي نحصل عليها من اختبار ستانفورد ـ بينيه أو اختبار وكسلر هي تقييم شامل لمستوى الذكاء. وقد بينت تحليلات معاملات الارتباط بين الاختبارات الفرعية أن هناك عاملاً عاماً كبيراً يتدخل فيها جميعاً. وهناك أدلة كثيرة من نتائج البحوث تؤيد ما ذهب إليه بينيه من أن الذكاء ليس مجموع كثير من القدرات الأبسط، ولكنه خاصة عامة تتأصل في الأنواع المختلفة من التفكير المعقد. ولازال من المعقول أن نفترض أن هناك شيئاً يسمى القدرة العقلية العامة، وإنه يمكن أن نقيسها عند شخص ما عن طريق مجموعة من الأسئلة والمهام. ولن يتعارض ذلك مع وجهة النظر التي تفترض أن هناك أيضاً مجموعة من القدرات الخاصة والمواهب، ولكن ربما لم يصمم كل من اختبار بينيه واختبار وكسلر لقياسها بكثير من الدقة.

اختبارات الذكاء الجمعية:

لقد تعرضنا في شيء من التفصيل لإثنين من أهم اختبارات الذكاء الفردية، وذلك لأن المفاهيم والقضايا الأساسية المتعلقة بها قد حدّدت بوضوح أكثر مما حدث بالنسبة للاختبارات الجمعية. ولكن بالنسبة لكثرة الاستخدام فإن الاختبارات الجمعية قد تقدمت كثيراً جداً عن الاختبارات الفردية، إذ أن ملايين من هذه الاختبارات الجمعية يطبق سنوياً في المدارس والمصانع والمنظمات الحربية. ولذلك، فإن على المواطن المثقف أن يعرف عن هذه الاختبارات القدر الكافي لكي يستطيع تفسير نتائجها بذكاء.

وعلى الرغم من أن اختبارات الذكاء الجمعية تشبه الاختبارات الفردية في نواح عديدة، إلا أن هناك أيضاً بعض الاختلافات الأساسية. فكل اختبار من هذه الاختبارات الجمعية يكون ذا غرض محدود بصورة أكثر من الاختبارات الفردية مثل اختبار بينيه أو وكسلر. فقد يكون الاختبار الجمعي مصمماً لمرحلة عمرية معينة أو مرحلة دراسية محددة، وبالتالي يصبح غير صالح لأطفال أصغر أو أكبر. ولذلك فإننا غالباً ما نجد الاختبارات الجمعية التي تغطي مراحل عمرية متتالية تطبع على هيئة سلسلة من الاختبارات. فعلى سبيل المثال هناك أربعة اختبارات هنمون ـ نلسون Henmon - Nelson Tests: أولها للسنوات التعليمية من ٣ ـ ٦، والثاني للسنوات التعليمية من ٦ ـ ٩، والثالث للسنوات التعليمية من ٩ ـ ١٢، والرابع للسنوات التعليمية من ١٣ ـ ١٧. والاختبارات

الجمعية تختلف أيضاً من حيث أنواع الأسئلة المستخدمة. فالاختبار المصمم خصيصاً لاختبار طلبة الجامعة مثل «اختبار الاستعداد المدرسي لدخول الكليات»[١] يختلف تماماً من حيث المحتوى عن اختبار جمعي آخر مصمم لتصنيف مجندي القوات المسلحة في برامج تدريبية مختلفة مثل «اختبار التصنيف العام للجيش»[٢].

ومثل هذا التخصيص يعتبر ميزة في مواقف معينة. وكقاعدة عامة، نجد أن الاختبار الجمعي الجيد يتنبأ بدرجات محكات خاصة بنجاح أكثر مما يقوم به الاختبار الفردي. فإذا أردنا أن نعرف مقدماً متوسطات درجات السنة الأولى للمتقدمين إلى كلية ما، فلن تكون هناك فائدة من إعطاء كل طالب اختبار وكسلر لذكاء الراشدين. وعلى الرغم من أن الدرجات على الجزء اللفظي من هذا الاختبار قد تتنبأ بدرجات الكلية تنبؤً لا بأس به، إلا أن أي اختبار من اختبارات الاستعدادات الجامعية الخاصة يمكنها أن تتنبأ بهذه الدرجات بشكل أفضل، بالإضافة إلى أن تطبيقه لا يستغرق إلا اليسير جداً من الوقت والنفقات إذا ما قورن بالاختبار الفردي.

والآن ما هي مميزات الاختبارات الفردية إذا ما قورنت بالاختبارات الجمعية؟ أولاً، إنها تمكّن الفاحص من التعرف بصورة أفضل على درجة دافعية المفحوص، كما تمكنه من تشجيعه ليبذل أقصى ما يستطيع من جهة في الإداء. فهناك احتمال قائم بأن درجة الفرد في اختبار جمعي قد تمثل تقديراً أقل جداً من قدرته، إذا كان لسبب أو لآخر لا يحاول أن يجيب عن أسئلة الاختبار إجابة صحيحة. وثانياً، إن الاختبار الفردي يعطي تصوراً أفضل للإمكانية العقلية لفرد ما قد تكون مهاراته في القراءة أقل من قدرته على التفكير والاستدلال. وعلى الرغم من وجود بعض الاختبارات الجمعية غير اللفظية، إلا أن أكثر هذه الاختبارات استخداماً يتأثر بمهارات القراءة، بحيث تعطي تصوراً غير صحيح لذكاء من تكون مهاراتهم في القراءة ضعيفة. وثالثاً، فإن الاختبارات الفردية تهيىء الفرصة للدراسة الكيفية والملاحظة. فالفاحص يستطيع أن يصف كيف يفكر الطفل المفحوص، كما يستطيع بعد تطبيق اختبار وكسلر أو اختبار

---

(١) The Scholastic Aptitude Test of the College Entrance Examination Boord.

(٢) The Army General Classification Test.

ستانفورد - بينيه أن يقترح بعض الأسباب المحتملة لنواحي النقص لديه بصورة أفضل مما لو استخدم اختباراً جمعياً.

وهناك أنواع عديدة من الاختبارات الجمعية يمكن لمن يعرفها جيداً أن يختار منها. فهناك اختبارات شديدة الصعوبة تتحدى تفكير طلبة الدراسات العليا، كما أن هناك ما هو كثير السهولة يناسب المتخلفين عقلياً. وهناك أيضاً اختبارات تعتمد كلية على اللغة، كما أن هناك اختبارات لا تستخدم فيها اللغة من أي نوع سواء في البنود ذاتها أو التعليمات. وهناك اختبارات تعطي فقط درجة كلية واحدة، وهناك اختبارات تعطي مجموعة من الدرجات الكلية المتمايزة. إن مناقشة كل هذه الأنواع من الاختبارات على حدة أمر يخرج عن موضوع اهتمامنا في هذا الكتاب. ولكن ما نريد أن نوضحه الآن هو أن جميع هذه الاختبارات تمكننا، بطريقة ما أو بأخرى، من أن نقارن القدرة العقلية العامة لفرد ما مع متوسط جماعة ما ينتمي إليها، أو جماعة ما يجب عليه أن يتنافس مع أفرادها.

اختبارات الأطفال الرضّع

هناك مجال معين من مجالات قياس الذكاء يستحق عناية خاصة ذلك لأن النتائج التي نحصل عليها فيه غير نموذجية تماماً. وهذا هو مجال اختبارات الأطفال الرضّع والأطفال الصغار جداً. فكل ما تعرضنا له من اختبارات قد صممت أساساً لأطفال في سن المدرسة وللراشدين. إن اختبار ستانفورد - بينيه الذي يمكن أن يصل إلى مستوى عمري أدنى من الاختبارات الأخرى، يتضمن مجموعة من الأسئلة التي تناسب الأطفال في عمر العامين، وقد استخدم كثيراً في اختبار أطفال ما قبل المدرسة، ولكنه لا يصلح لأطفال أقل من عامين، أو حتى لأغبياء الأطفال في سن العامين. وهناك، إذن، سؤال لا يمكن تجاهله وهو: هل يمكن تقدير ذكاء طفل لم يصل بعد إلى هذا العمر - أي عامين؟ فلو أمكن إيجاد مثل هذه المقاييس فإن ذلك سوف يساعد كثيراً في بعض الأغراض العملية مثل وضع الأطفال في بيوت محتضنة (أو بديلة)، كما سوف يساعد أيضاً في كثير من أغراض البحوث المختلفة.

لقد كانت هناك، في الواقع، عدة دراسات ذات تخطيط جيد عن الأشياء التي يمكن أن يقوم بها صغار الأطفال من مختلف الأعمار. وقد صممت عدة مقاييس للأطفال الرضّع تمكننا من معرفة مدى تقدم نمو الطفل. وعند إجراء

مثل هذا الاختبار فقد يقوم الفاحص، مثلاً، بتقديم حلقة لطفل عمره ثلاثة شهور ويلاحظ هل سيتمكن الطفل من الوصول إليها أم لا. وقد ينصت الفاحص إلى ما يصدره طفل عمره تسعة شهور من أصوات ليعرف ما إذا كان هذا الطفل يستطيع أن يقول با ـ با، أو شيئاً يقارب ذلك من حيث التركيب. وإذا كان هذا المثال الأخير يبدو إنه يقيِّم بداية نمو اللغة عند الطفل، فإن نسبة كبيرة من بنود مقاييس الأطفال الرضّع تعكس المهارات الحسية الحركية.

إن مثل هذه الاختبارات المقننة في نمو الأطفال الرضّع أصبحت ذات فائدة في قياس حالة الطفل في وقت تطبيق الاختبار عليه. كما إن هذه الاختبارات مفيدة للمتخصصين في طب الأطفال وغيرهم ممن يتعامل مع الأطفال الصغار جداً ومن هم في حاجة إلى أن تكون معالجتهم أو توصيتهم قائمة على أساس العمر النمائي للصغير وليس على عمره الزمني، حيث أنه في أي مرحلة عمرية نجد أن بعض الأطفال أكثر تقدماً عن البعض الآخر.

ولقد تراكم الدليل على أنه لا يمكن التنبؤ على أساس هذه الاختبارات بمدى سرعة نمو ذكاء الطفل فيما بعد. فالطفل الذي قد يكون بطيئاً في الجلوس بمفرده وفي الوصول إلى الأشياء على منضدة أمامه قد يكون أول من يتعلم القراءة في صفه أو أول من يجيد الخطوات المعقدة للقسمة المطولة. وباستثناء حالات التخلف العقلي الشديد، فإنه لا يمكن تقدير نسبة ذكاء الطفل عندما يكبر عن طريق أي اختبار يعطى له خلال السنة الأولى من عمره.

اتجاهات التقدم في قياس الذكاء:

يمكن وصف ما تقيسه اختبارات الذكاء بأنه القدرة على التعامل مع الرموز. والتفكير الذي عن طريقه تعالج الرموز لا الأشياء ذاتها، يعتبر مكوِّناً رئيسياً في الحياة المتمدينة. وكلما كان الإنسان أكثر ذكاء، كانت هذه الرموز أكثر تعقيداً وتجريداً. وتزداد قدرة الأطفال على التفكير الرمزي مع تقدمهم نحو النضوج، كما يميلون أن يصبحوا أكثر تخصصاً بالنسبة لأنواع الرموز التي يمكنهم التعامل معها بكفاءة أكثر. ولذلك فإن قياس هذه الإمكانية كان ذا قيمة كبيرة.

وبالرغم مما كان لاختبارات الذكاء من فائدة، إلا أنها لم تكن نعمة خالصة من أية شائبة. فمن ناحية، كانت بعض الآراء الخاطئة والقديمة حول ما تقيسه هذه الاختبارات تميل إلى الاستمرار وتعمل على إصدار أحكام غير

صحيحة عن الأفراد. فالذكاء كما تقيسه الاختبارات إنما هو سمة أكثر تحديداً مما يقوم الناس غالباً بتفسيره. ونسبة الذكاء ليست مؤشراً على الخاصة البشرية العامة. فهي لا توضح لنا كيف يكون الإنسان «موهوباً» في الفن، أو الموسيقى، أو الميكانيكا، أو حتى في التفاهم البشري. وهي لا توضح لنا إلى أي مدى سوف يتكيف الإنسان بنجاح للمواقف الجديدة. كما إنها لا تدل أيضاً على مدى سرعة وسهولة تعلم الإنسان للأشياء الجديدة في المواقف غير المدرسية التي سوف يجابهها. وأهم من هذا كله فإن نسبة الذكاء ليست مقياساً نقياً لقدرة فطرية، ولكنه يعكس الخبرة كما يعكس الإمكانية، ويعكس التربية كما يعكس الاستعداد.

إن الجهود المستمرة لتحسين ممارسة قياس الذكاء قد تقدمت في عدة اتجاهات. فنجد بعض الإخصائيين النفسيين قد اهتموا كثيراً بتفصيل ذلك المفهوم الشامل للذكاء إلى مجموعة أدق من المفاهيم تتعلق بأنواع خاصة من القدرة العقلية. وكلما أمعنوا في دراسة القدرات العقلية الأولية اكتشفوا عدداً كبيراً من هذه القدرات المتمايزة أكثر مما كانوا يتوقعون. فبحوث لويس ل. ثرستون Louis L. Thurstone في عام ١٩٣٠ قد أدت إلى اختبارات من الممكن استخدامها في قياس القدرة اللغوية، والقدرة المكانية، والسرعة الإدراكية، والذاكرة، والتفكير الاستدلالي، وأنواع أخرى متعددة من الذكاء. وإن أبحاث جيلفورد J. P. Guilford التي استمرت في هذا الاتجاه عدة سنوات منذ ذلك الوقت قد بينت أن هناك ١٢٠ نوعاً أو أكثر من القدرات العقلية المتمايزة. وقد وضعت اختبارات خاصة بكثير من هذه القدرات يمكن الحصول عليها، وعلى الأقل من أجل أغراض البحث العلمي.

وهناك اتجاه آخر في البحث تركز عليه اهتمام خاص وهو قياس الإبداع. وقد وضعت لهذا الغرض اختبارات ذات أسئلة ليست لها «إجابات صحيحة» محددة، وإنما تتطلب من المفحوص أن يفكر في بدائل متعددة من الإجابات، أو يعطي استجابات جديدة وأصيلة. فعلى سبيل المثال، قد يسأل المفحوص في أحد الاختبارات أن يكتب أكبر عدد يمكن أن يفكر فيه من الاستعمالات غير العادية لشيء ما مألوف، مثل علبة من الصفيح أو جريدة. وقد يسأل المفحوص في اختبار آخر أن يفكر في عناوين مناسبة لبعض الصور. وليس من الواضح تماماً أن الأطفال الذين يقومون بمثل هذه الأعمال جيداً سوف يكونون فيما بعد

مفكرين وفنانين مبدعين، ولكن هناك دليلاً معقولاً على أن هؤلاء الأطفال ـ نوع مختلف عن هؤلاء الذين يحصلون على أعلى الدرجات في اختبارات الذكاء العادية. ولذلك، فإنه يبدو أنه من المهم جداً أن نكتشف هؤلاء الأطفال، وأن نعرف الكثير عن نموهم.

وهناك اتجاه جديد آخر في قياس الذكاء، وهو اتجاه يعنى ببناء اختبارات يمكن أن يستخدمها المعلمون لإثارة النمو العقلي عند صغار الأطفال. فمن الأفكار ذات التأثير الواضح والتي قامت على مجموعة من البحوث تدور حول المراحل التي يمر خلالها الأطفال ونوعية الخبرات التي تمكنهم من الانتقال من مرحلة إلى أخرى، ما نجده في العديد من الكتب التي نشرها الإخصائي النفسي السويسري جين بياجيه Jean Piaget. وكانت هناك محاولات عديدة لبناء اختبار للذكاء يبنى على مفاهيم بياجيه. فقد أعدت مؤسسة خدمات القياس التربوي لمدارس مدينة نيويورك بعض المؤلفات من هذا النوع: «نظرة إلى تلاميذ المرحلة الأولى»، «التدريبات التجريدية لتلاميذ المرحلة الأولى»، «مواد التدريس والتقويم لتلاميذ المرحلة الأولى». فالمؤلّف الأول يعتبر مرشداً للمعلمين يوضح لهم ما هي أنواع الأشياء التي يبحثون عنها في الأطفال كدليل على النمو الذي يجب على المعلمين أن يحاولوا مساعدة الأطفال فيه. والشكل رقم ١٢ يوضح هذه الأشياء. وأما المؤلفان الآخران، مع المواد المصاحبة لهما، فيعطيان تدريبات عملية ووسائل للتقدير. وليست هناك معايير، أو درجات، أو نسب ذكاء على الرغم من أن كثيراً من هذه التدريبات تشبه تلك التي تستخدم في اختبارات الذكاء العادية. ويبين شكل ١٣ مثالاً لذلك. والخطة هي حث وإثارة كل طفل ليتقدم بالسرعة الممكنة في سلمه النمائي الخاص به، وليست مقارنة تحصيل هذا الطفل بتحصيل الأطفال الآخرين.

وعلى الرغم من أخطاء اختبارات الذكاء، فإنها أدوات لا يمكن الاستغناء عنها في المجتمع الحديث. فنحن نستخدمها لكي تساعدنا في اتخاذ قرارات حول وضع الأفراد في المدرسة أو العمل، ولكي تساعدنا في صياغة السياسات التربوية والاجتماعية. وكلما تقدمت البحوث في ميدان قياس الذكاء، أصبحت الاختبارات أكثر دقة وملاءمة. وهي، مثل كل الأدوات، تحتاج إلى مهارة خاصة في التناول، ومعرفة دقيقة لما يمكن أن تعمله، وما لا يمكن أن تعمله.

١ ـ مهارات أساسية في اللغة

- التميز السمعي والانتباه
- الفهم عن طريق الإنصات
- تعلم التواصل
- اللغة من أجل التفكير

٢ ـ مفاهيم المكان والزمان

- تعلم الأشكال والهيئات
- المنظور المكاني
- مفهوم الزمن

٣ ـ بداية المفاهيم المنطقية

- التطبيق المنطقي

مفاهيم العلاقات

٤ ـ بداية المفاهيم الرياضية

- بقاء الكمية
- طابق الأعداد
- العلاقات الرقمية

٥ ـ نمو مهارات التفكير الاستدلالي

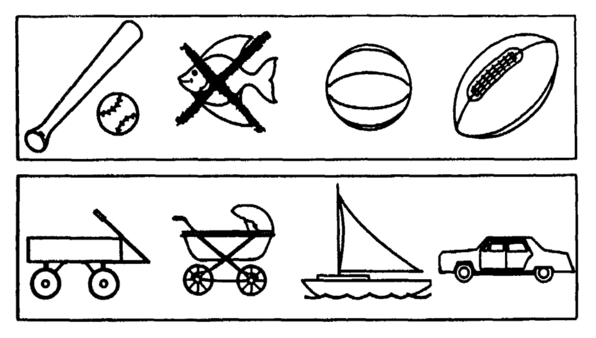
- فهم السبب والنتيجة
- الاستدلال عن طريق الترابط
- الاستدلال عن طريق الاستنتاج

٦ ـ علاقات عامة للنمو

- نمو الوعي وقابلية الاستجابة
- النشاط الموجه
- المعلومات العامة
- نمو التخيل

شكل ١٢ ـ ميادين النمو العقلي

(From **Let's Look at First Graders** Princeton, N. J.: Educational Testing Service, 1965).



شكل ١٣ ـ بنود من كراسة «التدريبات التحريرية لتلاميذ المرحلة الأولى». (برنستون، مؤسسة خدمات القياس التربوي، ١٩٦٥ (Princeton, N J.: Educational Testing Service, 1965). يطلب من الطفل استخراج الصورة التي لا تتمشى مع بقية الصور، ومن ثم يكون قد فهم المفهوم الذي تتضمنه الصور.

الفصل الخامس

اختبارات القدرات الخاصة

الاستعداد والتحصيل:

إن مجرد الملاحظة العابرة للناس تعتبر كافية لتوضيح أن الذكاء ليس هو كل شيء يلزم من أجل النجاح. هناك فرد ما «ب» يتفوق تفوقاً كبيراً في مقررات دراسة العلوم بالجامعة على الرغم من انه كان قد حصل على درجة متوسطة في اختبار الذكاء الذي أداه عند التحاقه بالجامعة، ولكنه كان قد درس العلوم والرياضيات بجدية وحماس منذ كان في الثانية عشرة، ولذلك فقد كوّن كمية كبيرة من المعارف النوعية وأجاد المهارات الأساسية لحل المشكلات. وشخص آخر «هـ» أصبح ميكانيكياً متميزاً لأن لديه قدرة غير عادية لمعرفة كيفية تناسب الأجزاء مع بعضها البعض. وثالث «جـ» عنده حساسية مدهشة للفروق الدقيقة في درجة النغم والإيقاع، وقد استفاد من هذه الموهبة من تعلم العزف على الكمان. وفتاة رابعة «ل» تعرف عن الشؤون المعاصرة والعلاقات الدولية أكثر بكثير مما يعرفه اي فرد آخر من صفها الدراسي بالمدرسة الثانوية، ولذلك أصبح من الطبيعي أن تكون مرشحة لتمثيل مدرستها في أحد المؤتمرات الدولية.

ويمكن أن نذكر عدداً لا حصر له من الأمثلة التي تبين حالات خاصة من التفوق.

ولقد اتخذ علماء النفس اتجاهين مختلفين لبناء اختبارات للتحقق من مثل هذه المواهب الخاصة: وهما قياس التحصيل وقياس الاستعدادات. وفي أول الأمر سلك هذان الاتجاهان طريقين مختلفين على نحوٍ ما، ولكن بعد ذلك تلاقى الاتجاهين في طريق واحد. ولا يزال من المعتاد بين الكتاب في ميدان القياس العقلي أن يميزوا بين اختبارات الاستعداد واختبارات التحصيل، غير أن التمييز قد أصبح من باب سهولة المعالجة أكثر منه بين مفاهيم أساسية.

ولا يزال مصطلحا الاستعداد والتحصيل يحملان معاني خاطئة نتيجة عوامل تاريخية. ففي فترات سابقة كان الاستعداد يعني عادة المواهب الخاصة التي كان من المسلم به أنها تعتمد على فروق فطرية أو وراثية بين الأفراد أكثر من اعتمادها على فروق تعود إلى الخبرة والتعلم. فإذا قلنا إن طفلاً ما عنده درجة عالية من الاستعداد الموسيقي، فإن ذلك يعني أن لديه ذلك النوع من الأذن والمخ الذي يساعده على تعلم المهارات الموسيقية المعقدة، ولكن لا يعني ذلك أنه حصل على بعض هذه المهارات من قبل. والذكاء، كما قيس بالاختبارات التي أشرنا إليها في الفصل السابق، كان يظن أنه نوع خاص من الاستعداد «الفطري» للعمل المدرسي. ففي كتاب قديم عن اختبارات الاستعدادت لمؤلفه جونسون أوكونور Johnson O'Connor كان عنوانه (وُلِد هكذا).

أما اختبارات التحصيل، فعلى العكس، قد كان يظن أنها تقيس ما تعلمه الإنسان. وكانت الحاجة إلى إيجاد اختبارات مدرسية أفضل من الأمور التي أثارت الاهتمام إلى البحث عن اختبارات تحصيلية أكثر ثباتاً وأكثر صدقاً - ففي كثير من المواقف المدرسية يصبح من المهم أن نحصل على تقدير دقيق لمقدار معلومات الطالب في الجبر، أو الأدب الإنجليزي، أو الكيمياء. وهناك أيضاً العديد من اختبارات التحصيل الخاصة صممت لمواقف غير مدرسية، مثل اختبارات الحرف التي تمكن من يقوم بمقابلة التوظيف من معرفة ما إذا كان الشخص الذي يدعي أنه عامل ماهر ما إذا كان ينتمي فعلاً إلى فئة العمال الماهرين.

وكما سبق أن أشرنا، فإن الاخصائيين النفسيين لم يعودوا يعتقدون إن

اختبارات الذكاء تقيس قدرة «فطرية» خالصة، وإنما تقيس خليطاً غير قابل للتحليل من الإمكانات الفطرية والخبرات التربويـة. وينطبق هـذا الاستنتاج كذلك على الأنواع المختلفة من اختبارات الاستعدادات. فالقدرة التي تقيسها اختبارات الاستعداد الميكانيكي، على سبيل المثال، إنما هي جزئياً نتاج الخبرة الميكانيكية. والقدرة التي يقيسها اختبار الاستعداد الكتابي إنما هي جزئياً نتاج الخبرات التي أدت إلى مهارة الفرد في إدراك التفاصيل الدقيقة، وكذلك القدرة التي تقيسها اختبارات الاستعـداد الموسيقي انمـا هي تعكس جزئيـاً التدريب الموسيقي . ففي الممارسة العملية نحن لا نستطيع أن نستخلص المكـونات «الطبيعية» من المكونات «المكتسبة» لأي استعداد من الاستعدادات، ولو أنه يمكن أن نفكر في كل منهما على حدة إذا شئنا ذلك.

وفضلاً عن ذلك، ففي الدراسات التي تجرى من أجل تصميم اختبارات للاستعدادت الخاصة بأنواع لأنواع من العمل والتدريب، نجد أن اختبارات التحصيل كثيراً ما تكون أفضل المؤشرات التنبئية للنجاح في هذا العمل أو هذا التدريب. وعلى ذلك، فإن الاختبارات التي تقيس المعلومات التي يعرفها الطالب عن المـوضوعـات التي تدرس في المـدارس الثانـوية ـ كـاللغة الانجليـزيـة، والرياضيات، والعلوم والتاريخ، وغير ذلك من المواد ـ إنما تستخدم بنجاح لتقدير الاستعداد للدراسة الجامعية. وقياس المعلومات عن الطيران كان جزءاً هاماً من بطارية اختبارات انتقاء الطيارين في القوات الجوية. واختبارات الهجاء كانت مفيدة في اختيار الموظفين الكتابيين.

ما هو، إذن، الفرق بين اختبارات الاستعدادات واختبارات التحصيل؟ فبالإضافة الى أن كلا النوعين له تاريخ يختلف عن الآخر، فإن الفرق الرئيسي يكمن في الغرض من الاختبار. فالاختبارات التي صممت واستخدمت أساساً من أجل اختيار (انتقاء) العاملين أو المتدربين تسمى اختبارات الاستعدادات. وأما الاختبارات التي تستخدم أساساً لمعرفة مدى ما تعلمه الطالب فتسمى اختبارات تحصيل. إن اصطلاح «القدرة»، وهو اصطلاح أعم، يشمل الاثنين معاً.

ومن الاختبارات التي نصنفها عادة على أنها مقاييس للاستعدادت تلك التي جاءت نتيجة برامج بحوث كبيرة ومشهورة. فهناك دراسة واسعة النطاق عن الاستعداد الميكانيكي في جامعة مينيسوتا أكملت في عام ١٩٣٠ وقام بها دونالد ج. باترسون Donald G. Paterson ومعاونوه، ونتج عنها ثلاثة اختبارات

ما زالت تستخدم بصفة مستمرة منذ هذا التاريخ حتى الآن. وهذه الاختبارات الثلاثة هي اختبار منيسوتا للعلاقات المكانية[١]، واختبار منيسوتا للتجميع الميكانيكي[٢]، ولوحة منيسوتا الورقية للأشكال[٣]. وفي نفس الوقت تقريباً، كانت هناك دراسة شاملة أخرى في جامعة آيوا انتجت مقاييس سيشور Seashore للمواهب الموسيقية[٤]. وهناك العديد من الأمثلة الأخرى، ولكن يكفي هذان الاختباران لتوضيح الأسباب التاريخية التي أدت عادة إلى وضع بعض الاختبارات في هذه الأيام في فئة اختبارات الاستعدادات. وقد أسهمت بعض الدراسات الأخرى كذلك في وضع اختبارات الاستعدادات الكتابية، وطرق تعيين المواهب الفنية، ومقاييس المهارة اليدوية والتآزر الحركي. وقد اعتبرت مثل هذه الاختبارات في الأغلب مقاييس استعدادات أكثر منها مقاييس تحصيل.

ولكن قد يكون أكثر فائدة لنا في الوقت الحالي أن نميز بين الاختبارات على أساس الغرض منها بدلاً من هذا التمييز القائم على أساس تاريخ البحث. فإذا كنا نريد أن نستخدم اختباراً ما أساساً لتنبأ بمدى نجاح بعض الأفراد في مجالات معينة، فيمكن أن نعتبره اختباراً للاستعدادات، بغض النظر عن كيفية التصنيف الذي وضعه من قام بإعداد الاختبار. أما إذا كنا نريد استخدام الاختبار أساساً لتقدير مدى إنجاز الفرد، أو كفاية تعليمه وخبرته، فيمكن أن نعتبره بالنسبة لغرضنا اختباراً في التحصيل، حتى ولو لم يكن مؤلفه قد فكر فيه على هذا النحو. واستخدام اختبار ما في أحد هذين الغرضين لا يمنع، بالطبع، من استخدامه أيضاً للغرض الآخر. فقد نريد أن نقدر تحصيل فرد معين ونتنبأ بنجاحه أيضاً وقد يكون نفس الاختبار مفيداً لكلا الغرضين.

الصدق من أجل أغراض مختلفة

إن ما يمتاز به هذا المدخل إلى القدرات الخاصة هو توضيح فكرتنا عن

---

(١) The Minnesota Spatial Relations Test.

(٢) The Minnesota Assembly Test.

(٣) The Minnesota Paper Form Board.

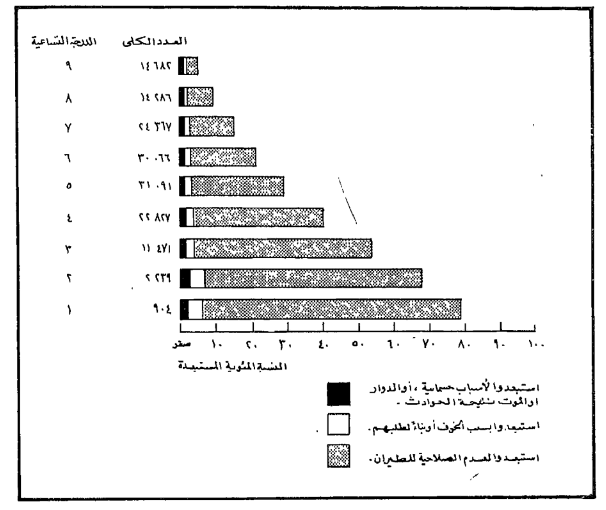
(٤) The Seashore Measures of Musical talents.

الأدلة التي يجب أن تتوفر لدينا لكي نبرهن على حقيقة صدق الاختبار. إن ما نهتم به هو صدق الاختبار من أجل أغراض معينة. فمن الواجب أن نتأكد من أن دليل الصدق الذي يقدمه مؤلف الاختبار له في الواقع علاقة وثيقة بهذه الأغراض، فإن مجرد ذكر معامل صدق من أي نوع لا يعتبر كافياً. فعلى سبيل المثال، لو أن مؤلف اختبار في الفهم الميكانيكي يقول إن معامل الارتباط بين درجات مجموعة من الصبيان في الصف التاسع على هذا الاختبار عندما طبق في شهر يناير وتقدير المعلمين لهؤلاء الصبيان في نفس الشهر هو ٠,٦٥، فإننا ما زلنا لا نعرف شيئاً عن الصدق التنبئي للاختبار. فإذا كان غرضنا هو اختبار الطلاب الصالحين للدراسة في مدرسة فنية عليا، فإننا نحتاج إلى دليل ما على أن هذا الاختبار قد سبق تطبيقه فعلاً على مجموعة مماثلة من الصبيان قبل التحاقهم بهذه المدرسة الفنية العليا، وأن معامل الارتباط بين درجاتهم على هذا الاختبار وتقديراتهم فيما بعد كان عالياً بدرجة معقولة.

وفي شكل ١٤ ترى حصيلة دراسة خاصة للصدق التنبئي لبطارية من الاختبارات استخدمت من أجل اختيار المتدربين في القوات الجوية خلال الحرب العالمية الثانية. والنتائج الموضحة في الشكل واضحة وغير غامضة، وبالتالي فإن استخدام هذه الاختبارات فيما بعد في اختيار الطيارين كان له ما يبرره دون شك. فالأرقام توضح أن قبول المتقدمين الحاصلين على درجات تساعيه عالية فقط سوف يقلل إلى أدنى حد نسبة الفقد في الأفراد خلال التدريبات الأولية. (الدرجة التساعية هي نوع من الدرجات المقننة، حيث الدرجة ٥ تمثل الدرجة المتوسطة، ١ تمثل الدرجة الدنيا، ٩ تمثل الدرجة العليا. انظر شكل ٨ ص ٦٣ الذي يبين الدرجة التساعية مع غيرها من الدرجات المشتقة الأخرى).

وإذا كنا نريد أن نستخدم اختباراً لقياس التحصيل بدلاً من الاستعداد، فإن نوعاً مختلفاً من أدلة الصدق يصبح ملائماً. إن السؤال الأساسي هو هل محتوى الاختبار يمثل موضوع المادة العلمية المطلوب قياس تحصيل الفرد فيها. ونحن نستطيع ان نجيب على ذلك بعد مراجعة أسئلة الاختبار. ولكننا نحتاج أيضاً إلى معرفة كيف اختيرت الأسئلة التي يحتويها الاختبار، وكذلك مَنْ الذي قام باختيارها. فعلى سبيل المثال، إذا أراد معلم في المدرسة الثانوية أن يستخدم اختباراً تحصيلياً مقنناً في نهاية تدريسه لمقرر من التاريخ الأمريكي، فإنه يحصل على نموذج من الاختبار الذي يبدو مناسباً، وكذلك كراسة التعليمات الخاصة

به. فإذا كان المعلم يعرف مؤلفي هذا الاختبار، أو على الأقل سمع بهم بسبب شهرتهم، ويثق في دقة أحكامهم على أهداف مقرر في التاريخ الأمريكي، فإنه سوف يبدأ باستخدام هذا الاختبار باتجاه طيب نحوه. وإذا ذُكر كذلك في كراسة التعليمات أن مجموعة من الخبراء قاموا بوضع التخطيط الأولي للاختبار وعملوا كمستشارين في مراجعة الأسئلة التي أثير حولها الشك، فإن ذلك سوف يؤكد لديه انطباعه الجيد. وإذا ذكر أيضاً أن مؤلفي هذا الاختبار قاموا بتجربته في صورته الأولية على مجموعة ممثلة من الطلاب للتأكد من أن معاني الأسئلة كانت واضحة وأن الأسئلة لم تكن شديدة الصعوبة أو كثيرة السهولة، فإن هذه تصبح نقطة أخرى في صالح هذا الاختبار.



شكل ١٤ - النسبة المئوية للمتقدمين الذين استبعدوا من التدريب الأولي للطيارين من حيث ارتباطها بالدرجات التساعية على بطارية الاختبارات التي استخدمت في الاختيار.

(From «Psychological Activities in the Training Command, Army Air Forces,» by the Staff, Psychological Section, Fort Worth, Texas, in **Psychological Bulletin** Washington, D. C.: American Psycological Association, Inc., 1945)

والخلاصة، إذن، هي أن اختبار القدرة يمكن اعتباره اختباراً في الاستعداد أو في التحصيل اعتماداً على الغرض الذي نهدف إليه من تطبيق الاختبار. فإذا كان سيستخدم كاختبار للاستعداد، وجب إثبات الصدق التنبئي. وإذا كان سيستخدم كاختبار للتحصيل، فإن صدق المحتوى يصبح هاماً.

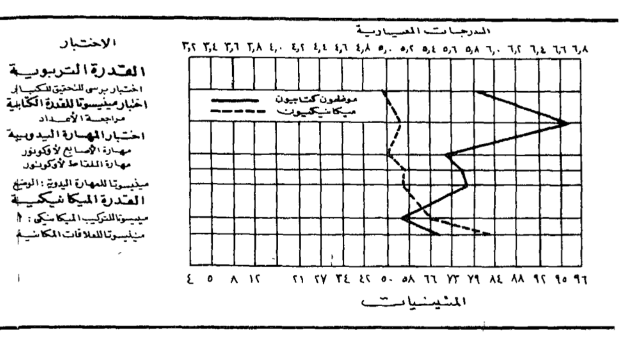
البحوث في ميدان القدرات الخاصة

الدراسات المهنية

لقد كان من بين الأهداف الرئيسية للرواد الأوائل في ميدان قياس الاستعدادات إيجاد الأدوات التي تساعد على التوجيه المهني القائم على أساس علمي. ولكن بمضي الوقت تلاشى هذا الأمل أو كاد، إذ أن الاستعدادات فيما يبدو، أكثر تعقيداً، وأكثر اعتماداً على أنواع خاصة من الخبرة السابقة، مما كنا نظن من قبل. فكثير من المواهب الخاصة ظهر أنه لا يمكن قياسها على الإطلاق، على الأقل بالطرق المعروفة حالياً. وكذلك فإن الاهتمامات والقدرات غالباً ما تختلف كثيراً حتى أن شخصاً ما قد لا يرغب في الالتحاق بمهنة معينة تتفق مع استعداداته التي حددتها وسائل القياس. وأخيراً، فإن التغير التكنولوجي قد أصبح سريعاً جداً بحيث إننا أصبحنا غير متأكدين من أن الاختبار ذا درجة الصدق التنبئي العالية لمهنة معينة في عام ١٩٧٠ سيكون له أي صلة على الإطلاق بما يزاوله الناس في عام ١٩٨٠.

ولكن هناك بعض الدليل الواضح على أن الأفراد الذين يقومون بمهن مختلفة يختلفون عن بعضهم البعض في قدراتهم الخاصة. فشكل ١٥ يبين الدرجة المتوسطة لبعض العمال في مجال الميكانيكا وبعض العاملين في الأعمال الكتابية في عدد من الاختبارات. ويتضح من هذا الشكل أن العاملين في الأعمال الكتابية أعلى كثيراً من عمال الميكانيكا من القدرة التربوية (ما يسمى عادة بالذكاء العام)، وفي القدرة الكتابية كما تقيسها اختبارات مراجعة الأرقام، وثلاثة أنواع من المهارات اليدوية؛ في حين ان عمال الميكانيكا أعلى من العاملين بالأعمال الكتابية في كلا الاختبارين اللذين يقيسان القدرة الميكانيكة. وعلى أساس مثل هذا الدليل يمكن للموجه المهني أن يوضح لطالب الاستشارة ما إذا كانت لديه نفس مجموعة القدرات التي لدى الآخرين الذين يمارسون العمل

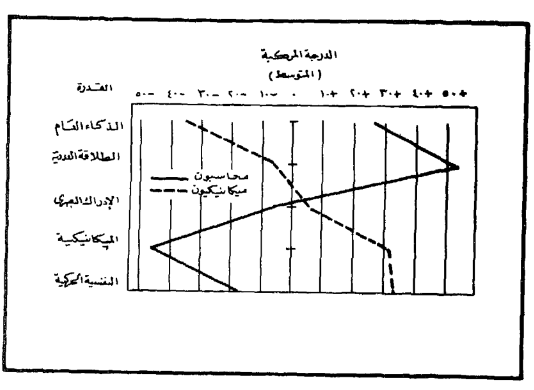
الـذي يرغب في الالتحاق به، حتى ولو لم يكن الموجه المهني قادراً على أن يتنبأ بدرجة نجاح هذا الفرد في ذلك العمل.



شكل ١٥ ـ بروفيل درجات العاملين في الأعمال الكتابية والمكانيكية.

(From B. J. Dvorak. **Differential Occupational Ability Patterns.** Employment Stabilization Research Institute, Vol. 3, No. 8. Minneapolis: University of Minnesota Press. Copyright 1935 by the University of Minnesota).

وهناك دراسة على نطاق واسع قام بها روبرت ل. **ثورندايك** Robert L. Thorndike **وإليزابت** ب. هاجان Elizabeth P. Hagan ـ **ذكـرت** في كتاب عشرة آلاف مهنة ـ تعطي دليلاً إضافياً يؤكد أن الجماعات المهنية تختلف من حيث نظام القدرات، وتوضح أيضاً ان درجة نجاح الفرد في مهنة ما لا يمكن التنبؤ بها من درجات الاختبار. وقد قام ثورندايك وهاجان في عام ١٩٥٥ بحصر عدد كبير من الرجال الذين طبقت عليهم بطارية اختبـارات القوات الجوية في فترة معينة من عام ١٩٤٣ (حوالي عشرة آلاف رجل)، وكانت الخطوة التالية هي البحث عن المهنة التي التحق بها كل رجل ـ وبعد ذلك قدر الباحثان درجة نجاح كل فرد في مهنته باستخدام عدة طرق ـ مثل المرتب الشهري، والترقية، والمؤهلات المهنية. وقد اختلفت بروفيلات الدرجات المتوسطة للمجموعات المهنية المختلفة بصورة واضحة، كما هو مبين في شكل ١٦. ولكن معاملات الارتباط بين تقديرات النجاح في المهن المختلفة وبين درجات الاختبارات كانت جميعها تقريباً قريبة من الصفر.



شكل ١٦ ـ بروفيل درجات المحاسبين والميكانيكيين.

(From R. L. Thorndike and E. Hagen. **10,000 Careers** • New York: John Wiley and Sons, Inc., 1959. pp. 27—28.)

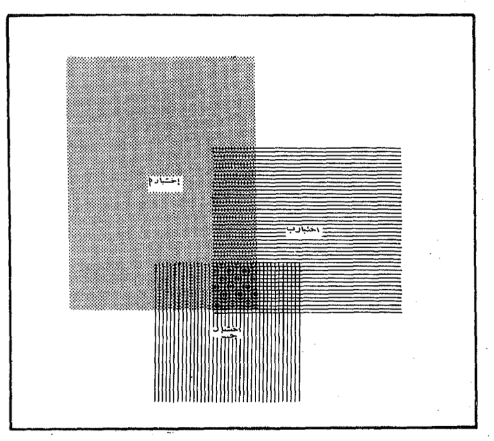
وعندما نطبق مثل هذه النتائج على الحالات الفردية يجب أن نتذكر أن الدرجات الموضحة في شكل ١٥، ١٦ ما هي الا متوسطات. والأفراد يختلفون كثيراً في نطاق المهنة الواحدة، حيث نجد أن البعض يفي بمتطلبات العمل بقدرة تقل كثيراً عن قدرة العامل المتوسط، والبعض الآخر تكون قدرته أعلى كثيراً. وحيث ان معظم الأعمال يمكن أن تتم بطرق عديدة، فإن من المحتمل أن يستخدم الأفراد استعدادات مختلفة للقيام بهذه الأعمال. ويعتبر ذلك صحيحاً بالنسبة للمهن التي تتطلب جهداً أكبر. ولكن من المحتمل أن هناك حدوداً لإمكان التبادل بين الاستعدادات، حيث أن العامل الضعيف جداً في أحد الاستعدادات الأساسية قد لا يمكنه القيام بعمل معين مهما أوتي من إمكانات مهنية أخرى. ولذلك، فإن التصنيف في عالم المهن يبدو أنه يتم بصورة عامة، على النحو الذي يعطينا النماذج التي نراها في شكلي ١٥ و ١٦.

التحليل العاملي

هناك تقدم آخر رئيسي في دراسة القدرات الخاصة وهو محاولة فهم ماهية هذه القدرات عن طريق تحليل الارتباطات بين درجات الاختبارات. وقد أت.

الى ذلك في الفصل السابق عندما ناقشنا الأدلة التي قدمها ثرستون وجيلفورد والآخرون على أن ما نقيسه كذكاء عام يمكن أن يحلل إلى أنواع من القدرات الفعلية أكثر تحديداً. وعندما تقدمت البحوث اتضح أن ما كان يسميه بعض الإخصائيين النفسيين أنواعاً من الذكـاء إنما هي نفس الخصـائص التي كان يسميها البعض، الآخر الاستعدادت المهنية. وكانت نتيجة ذلك بناء بطاريات من الاختبارات بنيت أساساً على التحليل العاملي ولكنها مرتبطة بمحكات مهنية.

إن التحليـل العامـلي منهج لتحليـل جداول الارتبـاطات بـين درجات الاختبارات من أجل توضيح العامل المشترك الذي يربط بين هذه الاختبارات المختلفـة. وبالـرجوع إلى شكـل ١٧ يمكن فهم المنطق وراء هـذه العملية.



شكل ١٧ ـ العلاقات بين ما تقيسه اختبارات مختلفة كما تفترضها نظرية التحليل العاملي.

لنفـرض أننا قمنـا بإجـراء ثلاثـة اختبارات عـلى مجموعـة كبيرة من تـلاميذ المدرسة الثانوية، وعن طريق الوسائل المشار إليها في الفصل السابق نحصل على ثلاثة معاملات ارتباط بين الدرجات على الاختبارات: أ، ب ـ أ، جـ ـ ب، جـ.

وقد كانت جميعاً موجبة، ولكن معامل الارتباط بين أ، حـ أقل قليلا من المعاملين الآخرين. فكيف نفسر ذلك إذن؟ نحن نفترض عادة أن هناك تداخلاً بين القدرات التي تقيسها هذه الاختبارات الثلاثة،وأن معاملات الارتباط توضح مدى هذا التداخل. ان المساحات المظللة في شكل ١٧ تبين ذلك. فالقدرات التي تمثلها هذه المساحات هي القدرات التي تؤثر على أداء الفرد في أكثر من اختبار واحد، وهي القدرات التي يبحث عنها القائم بالتحليل العاملي. وفي الحقيقة، ان الموقف الذي يتعامل معه هذا الباحث أكثر تعقيدا من ذلك الذي يوضحه شكل ١٧. فقد يبدأ باستخدام خمسين اختباراً بدلاً من هذه الثلاثة، وعلى ذلك فسوف يتعامل في وقت واحد مع ١٢٢٥ معاملاً من معاملات الارتباط ($\frac{٥٠ \times ٤٩}{٢ \times ١}$ =١٢٢٥). وبعد أن يجد أي هذه الاختبارات يجمعها عامل عام، فإنه يعطي لهذا العامل (إسمًا ما) بناء على تحليله لعملية التفكير، أو خلفية الخبرة، أو المهارة الخاصة التي تبدو مشتركة بين الاختبارات المتداخلة التي تحدد هذا العامل العام.

وعند استخدام الاختبارات التي صممت عن طريق التحليل العاملي، فإنه من المهم ألا نعول كثيراً على «حقيقة» القدرات التي يبدو أنها تقيسها. فالعوامل هي أساساً مجرد افتراضات حول الارتباطات البينية في بطارية من الاختبارات. وغالباً ما تكون هذه الافتراضات مفيدة جداً من التفكير حول الأفراد، ولكن من الضروري مراجعة هذه الافتراضات في مقابل معلومات أخرى قبل الوصول إلى استنتاجات نهائية.

بطارية الاستعدادات العامة

أن أشمل البرامج التي تجمع بين البحث في المهارات المهنية ودراسة التحليل العاملي للارتباطات بين درجات اختبارات الاستعدادات قد بدأته في الأربعينات المبكرة هيئة خدمة التوظيف الأمريكية. فقد أخذ الباحثون على عاتقهم مهمة تحديد القدرات الأساسية للأنواع المختلفة من المهن، ومن ثم بناء الاختبارات المناسبة لقياس هذه القدرات، وبالتالي تحديد حد أدنى من الدرجات على كل اختبار من هذه الاختبارات. فبطارية اختبارات الاستعدادت العامة[١] تقوم على كل من تحليل العمل الذي يؤدي إلى اختبارات ذات صدق تنبُّئي

(١) The General Aptitude Test Battery (G A T B).

والتحليل العاملي الذي يؤدي الى اختبارات في القدرات الأساسية. وتقيس هذه البطارية تسعة استعدادات على مدى ساعتين ونصف، وهذه الاستعدادات هي: الذكاء العام (G)، القدرة اللغوية (V)، القدرة العددية (N)، القدرة المكانية (S)، إدراك الشكل (F)، الإدراك الكتابي (Q)، التآزر الحركي (K)، مهارة الأصابع (F)، المهارة اليدوية (M). وقد وضعت درجات دنيا للاستعدادات الأساسية لكل نوع من المهن. ولحسن الحظ، فإن النظام الواحد من الاستعدادات الأساسية غالباً ما يميز مجموعة من الأعمال المتقاربة، ومن ثم فإن عملية التفسير ليست أمراً صعباً. وباستخدام بطارية الاستعدادات العامة يمكن للإخصائي الذي يقوم بالإرشاد المهني أن يقارن بروفيل درجات الفرد بكل من نماذج القدرات المهنية الستة والثلاثين التي أمكن تحديدها من خلال البحث. وهذه النماذج الأساسية الستة والثلاثون تغطي عدة مئات من الأعمال. وعلى هذا، فإن الفرد الذي تطبق عليه بطارية الاستعدادات العامة يمكن أن يلفت نظره لميادين أعمال كثيرة لم يفكر فيها من قبل على الإطلاق، ولكنه قد يكون صالحاً لها، كما أنه سوف يعرف أيضاً ما إذا كانت لديه قدرة خاصة معينة أقل من الحد الأدنى المطلوب لعمل ما كان يفكر في الالتحاق به. ويعتبر هذا، بالنسبة لشخص يفكر في مستقبله عائداً جيداً لاستثمار مدة ساعتين ونصف يقضيهما في الاختبار. فبطارية اختبارات الاستعدادات العامة توضح ما يمكن عمله حينما تُجمّع كل المعرفة التي تراكمت خلال سنوات عديدة عن الاستعدادات المهنية الخاصة حول مشكلة عملية معينة. ومع ذلك، فما زالت الحاجة قائمة لجمع أدلة أكثر عن الصدق، كما ان استنتاجاتنا عن شخص ما يجب أن تكون بحذر. وككل اختبارات الاستعدادات فإن هذه البطارية من الاختبارات ليست بعيدة عن الخطأ.

احتياطات في استخدام اختبارات الاستعدادات.

إن تفكيرنا الحالي عن طبيعة الاستعدادات الخاصة والمواهب وحول طرق بناء اختبارات لقياسها انما ينبع من هذين الاتجاهين في البحوث ـ الدراسات العملية في التنبؤ والتحليل النظري للذكاء العام عن طريق التحليل العاملي. والاختبارات المطبوعة والتي قامت على كلا الاتجاهين من البحوث أصبحت متوفرة للاستخدام في المدارس، ومكاتب المستخدمين والموظفين،

والخدمات الاستشارية، ولقد أصبح واضحاً أنه يجب أن نكون على حذر عند القيام باستنتاجات من مثل هذه الاختبارات. وعند فحص المتقدمين لعمل ما أو لبرنامج تدريبي، فإنه من الضروري عادة أن نجرب الاختبارات التي ننوي استخدامها في هذا الموقف الخاص، إذ أن الارتباطات بين الاختبارات والمحكات تختلف كثيراً من مكان إلى آخر. وعند استخدام اختبارات الاستعدادات لمساعدة الأفراد للوصول إلى قرارات تتصل بالمهنة، فإنه من المهم أيضاً أن نستخدم الاختبارات التي سبق أن جربت في بعض المواقف العملية حتى نكون على معرفة بعلاقاتها بمحكات الحياة الواقعية. إن اختبارات القدرات الخاصة التي تقوم فقط على التحليل العاملي للارتباطات بين الاختبارات لا تقول للفرد ما يحتاج أن يعرفه عن طريقه أدائه المحتمل في أحد مواقف المحكات. ولذلك فقد تؤدي إلى قرارات غير حكيمة.

وهناك خطأ ممكن آخر في استخدام اختبارات الاستعدادات جاء نتيجة الجهود التي بذلت من أجل توجيه «المعوقين» إلى الأعمال الإنتاجية. فقد يكون الاختبار متحيزاً ضد جماعات ذات خلفيات معينة غير عادية، وعلى ذلك فلن يعطي تنبؤات صادقة عن مدى نجاح أفراد هذه الجماعات في مهمة أو برامج تدريبية معينة. ولكي نعرف هل إن اختباراً معيناً متحيز أو غير متحيز على هذا النحو، فإن ذلك يحتاج إلى دراسة خاصة لهذه المشكلة. وكما أكدنا في الفصل الثالث فإن الموضوعية هي أحد المعايير الأساسية التي تقوّم بها الاخنبارات لأي غرض ما.

خصائص واستعمالات اختبارات التحصيل

لننظر الآن إلى عدة حقائق ومبادىء نحتاج أن نتذكرها دائماً عندما نستخدم اختبارات القدرات الخاصة التي صممت لقياس التحصيل - وهو ما تعلمه أو أنجزه الفرد.

وأساساً، إن الاختبار التحصيلي المدرسي المقنن ما هو إلا صورة منقحة من الاختبار الذي يستخدمه المعلم في نهاية تدريسه للمقرر الدراسي. إن مثل هذه الاختبارات قد صممت لتقيس بدقة وعلى نطاق واسع ما يقيسه كل معلم بانتظام عندما يريد أن يعرف إلى أي مدى تمكن كل طالب من المادة التدريسية. وعلى ذلك فليس من قبيل المصادفة أن تنتشر الاختبارات التحصيلية في الولايات

المتحدة الأمريكية انتشاراً كبيراً. فنظامنا في التعليم العام الشامل في المرحلتين الابتدائية والثانوية في أمريكا يحتم على المعلمين ومديري المدارس أن يتعاملوا مع طلاب على درجة كبيرة من التباين من حيث الخلفية الأسرية والاجتماعية. ولأن هذه المدارس تدار بنظام محلي (لامركزي) فليس هناك منهج موحد لجميع المدارس على مستوى الدولة، وكذلك نجد أن المدارس في بعض المناطق أكثر فعالية من مدارس في مناطق أخرى. ولأن الأمريكيين يغيرون مكان إقامتهم من حين لآخر، فإن الأطفال ينتقلون باستمرار من نظام تدريسي إلى نظام تدرسي آخر. ومن الواضح أنه أصبح من الضروري إيجاد طريقة مقننة تمكننا من معرفة مقدار ما يعرفه كل طفل من مادة دراسية معينة حتى يمكن أن نضمن سلامة النظام التربوي. وعلى ذلك، فإن اختبارات التحصيل التربوية تفي بحاجات اجتماعية حقيقية.

إن معظم طرق البحث التي ناقشناها سابقاً ساعدت على بناء بطاريات من الاختبارات التحصيلية مثل اختبار آيوا للنمو التربوي(١)، والاختبارات المتتابعة للتقدم التربوي(٢)، واختبارات القبول الجامعي(٣)، واختبار سجل التخرج(٤). ولكن عملية بناء هذه الاختبارات في حد ذاتها عملية ذات سمات مميزة. أولها وأهمها التأكيد على التحديد الدقيق لنوع المعلومات المطلوب اختبارها قبل البدء في بناء الاختبار. وهذه عملية ليست سهلة كما تبدو لأول وهلة. فإذا كان الفاحص مسؤولاً - على سبيل المثال - عما يجب أن يحتويه اختبار في التاريخ الأمريكي لطلبة المدرسة الثانوية، فإنه سوف يواجه بعض المسائل الصعبة: فهل يجب أن يؤلف الاختبار من أسئلة واقعية عن الأسماء والأماكن والتواريخ؟ فهذه أسهل في التأليف وفي التصحيح، ولكن معظم مدرسي التاريخ لا يعتبرون نقل مثل هذه المعرفة الواقعية الهدف الرئيسي من مقرر في التاريخ. أو هل من الواجب أن تكون الأسئلة حول الحركات والتيارات التاريخية؟ وإذا كان الأمر كذلك فأي حركات؟ وأي تيارات؟ والمؤرخون يختلفون في درجة تأكيدهم على الحركات السياسية والاقتصادية والثقافية، وكذلك تختلف الكتب ويختلف

---

(١) Towa tests of Educutional Development (ITED)

(٢) Sequential Tests of Educational Progress (STEP).

(٣) College Entrance Examination Board Tests(CEEB).

(٤) Graduare Record Examination (GRE).

المعلمون. ثم هل من الواجب أن يقيس اختبار في التاريخ الأمريكي المهارات والمعارف التي تهتم المقررات في هذه المادة بتعليمها؟ وهل يجب أن نضمِّـن مثل هذا الاختبار الأسئلة التي تساعدنا على الحكم على مدى قدرة الطالب على قراءة وفهم النصوص التاريخية، او الحكم على مدى قدرته على التفكير عندما يتتبع الجذور التاريخية لمشكلات قائمة في عصرنا الحاضر؟.

وكما اقترحنا سابقاً فإن أفضل طريقة لمعالجة الخطوة الأولى في بناء اختبار في التحصيل التربوي هي أن تشكل لجنة من المعلمين الأكفاء لمناقشة المشاكل المطروحة والوصول إلى قرارات حول ما يجب قياسه. وحتى بالرغم من أن مثل هؤلاء الخبراء يمكن أن يختلفوا فيما بينهم حول بعض التفاصيل، إلا أنهم في العادة يستطيعون الوصـول إلى اتفاق حـول المحتويـات والأهداف الأسـاسية للمقرر. وهذه غالباً ما تكون محور المعارف والمهارات الأساسية التي يجب أن يبنى عليها الاختبار حتى يصبح مناسباً لأكثر عدد ممكن من الفصول والمدارس.

وبمجرد تحديد هذا المحور والتخطيط الهيكلي المناسب له، تصبح الخطوة التالية هي وضع بنود الاختبار التي يجب أن تكون واضحة وموضوعية وذات معنى. وحيث أن هناك الآلاف من الطلاب سوف يؤدون هذا الاختبار وسوف تصحح أوراقهم آلياً، فلا بد أن يكون لكل بند من هذه البنود إجابة صحيحة واحدة فقط. وفضلاً عن ذلك، فإن هذه الاجابة يجب ألا تكون سهلة أو واضحة جداً حتى يمكن للاختبار أن يحقق في الواقع هدفه وهو التمييز بين الطلاب في مستوياتهم المختلفة من حيث المعرفة والمهارة. ان كتابة البند الجيد فن. حقيقةً إن هناك قواعد لما يجب أن نعمله وما لايجب عمله، ولكن المهارة، مثل الأنواع الأخرى من المهارة الكتابية، ليست في الواقع قابلة للتحليل.

وعندما يتم إعداد البنود وتجمع معاً في صورة قابلة للتجـريب، فإنـه يمكن استخدام الوسائل الإحصائية المناسبة لمساعدة مؤلف الاختبـار على أن يقـرر أي هذه البنـود يجب أن يبقى في الصورة النهـائيـة لـلاختبـار. وهـو يقـوم أولاً، بتهيئـة بضـع مئـات من الأفـراد منـاظـرين في العمـر، أو الصف الدراسي لهؤلاء الذين يؤلف من أجلهم الاختبار لكي يطبق عليهم الصورة التجريبية من الاختبار، ثم يرتب إجاباتهم على كل بند في جداول. ومن هذه الجداول يمكن أن يحسب معاملات الصعوبة والتمييز. أما بالنسبة للصعوبة، فمع أنه من المفضل عادة أن يحتوي الاختبار مـزيجاً من الأسئلة

الصعبة والسهلة، إلا أن الأسئلة السهلة جداً بحيث أن كل فرد تقريباً يجيب عليها إجابة صحيحة، أو الأسئلة كثيرة الصعوبة بحيث أن كل فرد تقريباً لا يجيب عليها، فإنها تكون عديمة الجدوى، وعلى ذلك فإن مثل هذه الأسئلة يحذف من الصورة النهائية للاختبار. أما عن معامل التمييز بالنسبة للبند فإنه يوضح ما إذا كان الطلبة ذوو المستوى العالي من نوع المعرفة التي يصمم الاختبار لقياسها (بناء على الدرجة الكلية العالية) أكثر قدرة من الطلبة ضعاف المستوى بالنسبة للإجابة على هذا البند. والبند الذي لا يميز على هذه الصورة فإنه عديم الفائدة في الاختبار، ويجب أن يحذف كذلك. وليس من الأمور غيرالمألوفة في مرحلة تحليل البنود لاختبار متعدد الاختيار أن نجد سؤالاً ما قد يختار إحدى الإجابات الخاطئة فيه الطلبة ذوو المستوى العالي أكثر من الطلبة ذوي المستوى المتدني. قد يحدث ذلك أحياناً لأن الطلبة ذوي المستوى العالي يستخدمون أسلوباً في التفكير أكثر تعقيداً ودقة مما توقعه مؤلف الاختبار. وعليه فإن مثل هذه الاسئلة لا بد أن تستبعد كذلك من الصورة النهائية للاختبار.

وبمجرد أن يتم إعداد مجموعة من البنود التي تفي بالغرض، فإن بقية خطوات بناء الاختبار التحصيلي تشبه كثيراً ما أشرنا إليه في الفصل الثالث. حيث أنه يتعين على مصمم الاختبار أن يبحث في ثبات الاختبار، ويعبر عنه كمعامل ثبات أو كخطأ معياري للمقياس. كما يجب عليه أن يحدد الطرق المقننة لإجراء الاخنبار وتصحيحه. ويجب عليه ان يعدّ درجات مشتقة ذات معنى من عينة ممثلة للمجتمع الذي يؤلف الاختبار له. ويجب عليه أيضاً أن يعدّ جداول المعايير. وكما سبق أن أوضحنا، فإنه إذا توفرت القرارات السليمة حول ما يجب أن يحتويه الاختبار، فإنه يمكن أن نطمئن إلى صدق المحتوى.

لقد تناولنا حتى الآن اختبارات التحصيل من وجهة نظر منتجها، ولكن ماذا يجب أن يتذكره المستهلك ـ المتعلم، أو مدير المدرسة، أو المفحوص ذاته ـ حينما يقوم بتقييم هذه الاختبارات والدرجات التي يحصل عليها منها؟. هناك عدد كبير من اختبارات التحصيل المتوفرة لقياس المعلومات من كل المواد المدرسية تقريباً وفي كل المستويات التربوية لدرجة أن انتقاء اختبار ما يصبح في كثير من الحالات عملية صعبة. ولهذا فإن ما يجب أن نهتم به أولاً هو ما أشرنا إليه سابقاً، وهو صدق المحتوى. فإذا أراد مدير المدرسة مثلاً، اختيار بطارية اختبارات تحصيلية لاستخدامها، فإن عليه أن يفحص الاختبار جيداً ويقرأ بدقة

الجزء الخاص من دليل الاختبار الذي يوضح كيفية تكوين الهيكل الأساسي للاختبار وكيفية اختيار الأسئلة أو البنود. فإذا وجد أن هذه العملية قد تمت بعناية ودقة فإنه يمكن أن يرى بصفة مبدئية احتمال استخدام الاختبار. ومن ثم ينتقل إلى فحص الخصائص الأخرى للاختبار - مثل درجة الثبات، وإمكانية التطبيق والتصحيح، وخصائص جماعات المعايير (أي التي اشتقت منها المعايير). ولعل ما يجب أن نؤكد عليه كثيراً هنا هو أن الحكم على صدق محتوى الاختبار يتطلب ما هو أكثر من فحص الأسئلة ذاتها. فإن ما يبدو على أنه سؤال يحتاج إلى إجابة بسيطة عن بعض الحقائق قد يحتاج إلى عمليات معقدة من التفكير. وإن ما يبدو على أنه مجموعة متوازنة من الأسئلة حول بعض المبادىء والمفاهيم المختلفة قد يكون غير متوازن بأن يميل أكثر إلى اتجاه معين. إن صدق المحتوى يعتمد على كل العمليات التي استخدمت في تخطيط وبناء الاختبار، أما الصدق الظاهري وحده - وهو الحكم الذي نصل إليه من مجرد النظر إلى الأسئلة - فهو لا يشكل أساساً سليمًا لانتقاء الاختبار.

بطاريات الاختبار الجامعة للاستعداد والتحصيل

حيث أنه لا يوجد خط واضح بين اختبارات التحصيل واختبارات الاستعدادت، ولأن اختبارات القدرات الخاصة المتنوعة تستخدم في المدارس ومكاتب الاستشارات، فقد تم تأليف بطاريات اختبارات جامعة لاستخدامها في الأغراض (المواقف) التربوية. من هذا النوع مجموعة استخدمت على نطاق واسع في المدارس الثانوية تسمى اختبارات الاستعدادات الفارقة[1]، وهي مكونة من ثمانية اختبارات منفصلة: (١) التفكير اللغوي، (٢) القدرة العددية، (٣) التفكير التجريدي، (٤) التفكير المكاني، (٥) التفكير الميكانيكي، (٦) السرعة والدقة الكتابية، (٧) الهجاء، (٨) الجمل. ونلاحظ أن القدرات التي تقيسها بعض هذه الاختبارات تتأثر إلى حد كبير بالخبرة المدرسية للطالب، في حين أن البعض الآخر يتأثر أيضاً إلى حد كبير بالخبرة خارج المدرسة. وقد قام مؤلفو هذه البطارية وناشرها بجمع كمية كبيرة من المعلومات حول ما يمكن التنبؤ به من درجات الطلاب على هذه الاختبارت. وقد لا تكون مثل هذه

---

(١) Differantial Aptitude Tests (DAT).

التنبؤات دقيقة وواضحة بالدرجة التي قد نتمناها ـ فليس من المؤكد، على سبيل المثال، أن الطالب الذي حصل على درجة عالية في القدرة العددية سوف يتفوق في مقررات الرياضيات على طالب آخر في فصله حصل على درجة عالية في التفكير اللغوي. ولكن عند دراسة المعلومات المصاحبة لهذا الاختبار يمكن للمعلم أو الإخصائي الذي يقوم بالإرشاد أن يحدد أي هذه التنبؤات ممكنة وإلى أي مدى يمكن أن تصاغ بدقة. وتتميز بطارية الاختبارات من هذا النوع على مقياس واحد للذكاء العام بأنه من المحتمل جداً أن تساعد هذه البطارية الطالب على اكتشاف بعض الإمكانات التي يمكن أن يخطط لمستقبله بناء عليها. كما أنها تتميز على مجموعة من الاختبارات المتفرقة للاستعدادات والتحصيل بأن معايير جميع اختبارات البطارية قد اشتقت من مجموعة واحدة ممثّلة. وبالتالي، فإن الدرجات التي توضح موقف الفرد بالنسبة للجماعة المعيارية يمكن مقارنتها من اختبار إلى آخر، بحيث يمكن أن نقول بكل ثقة، مثلاً، إن مستوى الفرد (أ) في التفكير الميكانيكي أعلى من مستواه في التفكير اللغوي.

وهناك أنواع أخرى عديدة من هذا النوع من بطاريات الاختبارات الجامعة، ومن المحتمل جداً أن تصبح أكثر انتشاراً بمرور الزمن. وطالما أن البحث في موضوع القدرات الفطرية قد انصرف عنه الاهتمام الآن، فقد تركز الاهتمام على سؤال آخر: وهو «ما هي الخصائص التي يكون من المفيد أن نقيسها في الفرد من أجل تسهيل عملية اتخاذ القرارات ـ تلك التي يتخذها الآخرون عنه أو التي يتخذها هو عن نفسه عن حياته الخاصة؟. ففي هذا الاتجاه يصبح من المفيد ان يعرف الفرد أي نوع من القدرات الخاصة يتميز بها، بغض النظر عن مدى نمو هذه القدرة وتقدمها. فمفهوم القدرة الخاصة، الذي يؤكد على ما يجب عمله بالنسبة للمواهب أكثر مما يؤكد على أصل هذه المواهب، قد يحل تدريجياً محل مفهومي الاستعداد والتحصيل اللذين لعبا دوراً هاماً في تطوير طرق القياس فيما مضى.

ومن البحوث التي تجرى حالياً والتي يمكن أن تلقي ضوءاً كافياً على الأسئلة التي تدور حول استخدام القدرات الخاصة، بحث يسمى «الموهبة» TALENT[1]. ففي ربيع ١٩٦٠، طبقت الاختبارات على ٤٤٠ ألف طالب في

(١) مشروع الموهبة TALENT تموله هيئة الصحة والتربية والرعاية، وتقوم به معاهد البحوث الأمريكية بالاشتراك مع جامعة بتسبرغ Pittsburgh.

١٣٥٣ مدرسة ثانوية على مدى يومين. وتضم بطارية «TALENT» مجموعة من الاختبارات التي صممت بدقة ورتبت بطريقة بحيث تقيس أكبر عدد ممكن من القدرات. ويحصل كل مفحوص على ٣٧ درجة مستقلة. وقد تم اختيار مدارس العينة بدقة لتمثل قطاعاً عرضياً كاملاً للمدارس الثانوية في الولايات المتحدة الامريكية. وبالإضافة الى درجات الاختبار، جمعت معلومات كافية ومتنوعة عن الطلاب والمدارس والمجتمعات المحلية. وكانت الخطة هي متابعة المفحوصين في السنوات ١٩٦١، ١٩٦٥، ١٩٧٠، ١٩٨٠ لمعرفة ما حدث لهم فيما يختص بالمهن التي اختاروها، والتعليم الذي حصلوا عليه، وانجازاتهم وإسهاماتهم في المجتمع. وعندما يتم تحليل هذا الكم الهائل من البيانات يمكن أن نعرف الكثير عما يحدث للأفراد ذوي المواهب الخاصة المختلفة تحت ظروف مختلفة. وفي كل مرحلة من مراحل المشروع، تنشر التقارير لإعلام الناس بما تم الوصول اليه.

وقد صاحب التقدم الذي حدث في اختبارات القدرات الخاصة ازدياد وعي الأمريكيين بأن المجتمع ككل سوف يستفيد عندما يشجع الأفراد على استخدام مواهبهم بفعالية وتحصيل كل ما يمكن تحصيله. وبناء على ذلك، فإن ما هو جيد بالنسبة للفرد (أ) هو جيد ايضاً بالنسبة للبلد. وطالما أن مثل هذه الاختبارات تستطيع أن تفيدنا في تحقيق الفائدة المشتركة لكل من الفرد والمجتمع، فإنها تستحق كل الجهد والحذق اللذين يبذلان في إعدادها، وبناء على ذلك، فهي تحتل مكاناً مرموقاً في البناء الكلي للقياس النفسي في الوقت الحالي.

الفصل السادس

تقدير الشخصية

عندما ننظر فيما حولنا إلى الناس الذين نعرفهم ـ بل وعندمانفحص أنفسنا نحن ـ فسوف نجد أن هناك بعض الخصائص الشخصية المميزة لا يمكن أبداً أن تصنف على أنها قدرات على الإطلاق. فنحن في العادة لا نعرف أو لا نريد أن نعرف شيئاً عن درجة جارنا على اختبار في القدرة الكتابية، ولكننا نحاول أن نقدر إلى حد ما درجة صداقته للناس، وحسن سلوكه، وتعاونه، وكرمه، واعتماده على الغير، واتجاهاته نحو الأمور السياسية. إن مثل هذه الخصائص ضرورية لفن الحياة، لأن النجاح يعتمد عليها تماماً كما يعتمد على المقدرة. ثم إن المشكلات الاجتماعية مثل الجريمة والمرض العقلي تتضمن وجود عادات شخصية معيبة في الأفراد أكثر مما تتضمن نقصاً في القدرة. ومن الواضح أيضاً أن النضج الانفعالي له على الأقل من الأهمية مثل ما للنضج العقلي من أهمية لنمو الأطفال من مرحلة المهد إلى مرحلة الرشد. وعلى ذلك، فليس غريباً أن يقبل الإخصائيون النفسيون التحدي لقياس سمات الشخصية كما هو الحال بالنسبة للقدرات.

صعوبات خاصة وطرق التغلب عليها:

إن قياس خصائص الشخصية يتضمن بعض المشكلات الخاصة بالإضافة إلـى تـلك التي طرحت في حالة قياس القدرات. إن أصعب هذه المشكلات، والتي تمثل الفرق الرئيسي بين اختبارات القدرات واختبارات الشخصية هي أن وضع المفحوص في موقف مقنن والحصول على عينة حقيقية من سلوكه هو أمر غير ملائم. فنحن بكل بساطة لا يمكن أن نجهز في غرفة الاختبارات المواقف المقننة التي يمكن أن تظهر فيها سمات الشخصية. فكثير من هذه السمات الهامة لها صبغة اجتماعية، وهي تظهر فقط عندما يجد الفرد نفسه في مجموعة معينة. فعلى سبيل المثال، إذا أردنا أن ندرس عينة لسمة ما مثل سمة ضبط النفس، فيجب أن نهيىء المواقف المقننة التي تسبب الإثارة، والإحباط، والمضايقة. (حاول علماء النفس أثناء الحرب العالمية الثانية القيام بذلك عن طريق مواقف توجه فيها الإهانة إلى أشخاص يحاولون التركيز على أعمال نفسية حركية معقدة، ولكن جهودهم لم تنجح. فتوجيه الإهانة من الفاحص يختلف عن توجيه الإهانة من رئيس العمل، حتى ولو ظهر أن الموقفين متشابهان). وفي حالة قياس القدرة، فإن استخدام عينة المواقف أمر ملائم. فالمشكلات التي تطرح في غرفة الاختبار يمكن أن تكون مشابهة في معظم الخصائص الأساسية للمشكلات التي يقابلها الفرد في المواقف الخاصة في الحياة. أما في قياس الشخصية فإنه من الصعب تشكيل المواقف الاجتماعية التي تتضح فيها الشخصية على وجه خاص. وقد كان هناك بعض البحوث الجادة التي حاولت ذلك، ولكن حتى الآن لم تتوفر منها اختبارات يمكن استخدامها على نطاق واسع.

إن المهتمين بقياس الشخصية قد تبنوا نوعين من الاستراتيجية للتغلب على هذه الصعوبة: أولهما الاستعانة بتقارير عن سلوك الإنسان بدلاً من ملاحظة هذا السلوك. ومن هنا نشأ العديد من استخبارات واستبيانات الشخصية التي تستخدم بكثرة حالياً. والثاني هو أن يعرض على المفحوص مثير غامض مثل بقعة حبر أو صورة ما ويطلب منه أن يستجيب لهذا المثير، وذلك على افتراض أن استجابته لا بد أن تعكس أسلوبه الشخصي الخاص وطريقة اتجاهه الخاص في الحياة. وسوف نعرض كلاً من هذين الأسلوبين على حدة.

استبيانات الشخصية

حينما يتبع مؤلف الاختبار الأسلوب الأول من الأسلوبين السابقين، وهو الاستعانة بالتقارير عن السلوك بدلاً من ملاحظته، فإنه يقوم أولاً بالتفكير في بعض الأسئلة أو العبارات التي تمثل مظاهر مختلفة للسمة التي يريد قياسها. لنفرض على سبيل المثال، أن هذه السمة هي **ضبط النفس**، فإن مؤلف الاختبار قد يفكر في إعداد عشرة أسئلة ذات علاقة بضبط الفرد لانفعاله عندما يتعرض لمواقف الإثارة أو الاستفزاز. وهو يفترض أن الفرد الذي يجيب بكلمة «نعم» على تسعة من هذه الأسئلة إنما هو على درجة من ضبط النفس أعلى من الفرد الذي يجيب بكلمة «نعم» على خمسة أسئلة فقط.

وقد ابتدعت هذه الطريقة في اختبار الشخصية في نفس الوقت تقريباً الذي ابتدعت فيه الاختبارات الجمعية في الذكاء. ففي أثناء الحرب العالمية الأولى قام روبرت س. وودوورث Robert S Woodworth بوضع مجموعة من الأسئلة أشبه ما تكون بتلك التي يسألها الأطباء النفسيون لفحص المجندين فيما يتعلق بالثبات الانفعالي. هذه المجموعة الأولى من الأسئلة والتي سميت صحيفة البيانات الشخصية[1]، كانت الأصل لمجموعة كبيرة من الاختبارات التي ظهرت فيما بعد في مجال التوافق العام والعصابية والسمات الأخرى المرتبطة بالمرض العقلي والصحة العقلية.

وهذه الطريقة - طريقة سؤال المفحوص لأن يجيب على أسئلة تتصل بسلوكه وإحساساته الوجدانية في المواقف الحياتية - تحمل في طياتها مشكلات تختلف عن تلك التي واجهناها في قياس القدرات. وفي الواقع، ان تاريخ قياس الشخصية يمكن أن ينظر إليه على أنه سلسلة من الجهود المبذولة لحل هذه المشكلات الخاصة. ويمكن ان نقول ان أوضح هذه المشكلات - وإن لم تكن أخطرها - هي أن الفرد لن يجيب بصدق عن نفسه. ونحن لا نشعر بقلق حول هذا الأمر في قياس القدرات، إذ ليس من المحتمل أن طفلاً ما له نسبة ذكاء قدرها ٩٠ يستطيع أن يزور لنفسه نسبة ذكاء قدرها ١٢٠. فإن لم تكن لديه القدرة العقلية التي تعمل على مستوى نسبة الذكاء ١٢٠، فإنه لن يستطيع الإجابة على الأسئلة التي تصل به إلى هذه الدرجة. ولكن من الممكن جداً

(١) Personal Data Sheet.

يقوم شخص شديد الخوف بإعطاء إجابات تدل على «الشجاعه» عند إجابته عن أسئلة حول استجاباته تجاه مواقف الخطر، كما أنه من الممكن للشخص المنطوي أن يجيب «بنعم» على سؤال يقول: «هل لك الكثير من الأصدقاء»؟ فإذا كان مدى ملاءمة الفرد لوظيفة ما، أو للقبول في كلية ما، أو للانضمام إلى نادٍ خاص يتوقف على درجته على اختبار في الشخصية، فإنه يقع تحت إغراء حقيقي لكي «يلفق الإجابة الأحسن».

وقد وجد الإخصائيون في علم النفس العسكري أيضاً أن الفرد قد يقوم في بعض الأوقات «بتلفيق الإجابة الأسوأ». فقد وجد أن بعض المجندين يظهرون أعراض الأمراض النفسية الشديدة عند الفحص المبدئي، ولكن يظهر انهم أصحاء تماماً عند إجراء بعض طرق الفحص الأكثر دقة. لقد كانوا يأملون أن يصنفوا كعصابيين، ولذلك أجابوا على الأسئلة بهذا القصد.

لقد كان الإخصائيون النفسيون بارعين جداً في مجال تصميم الوسائل لاكتشاف ما إذا كانت الإجابة ملفقة، بحيث يمكن إلغاء الدرجه أو تصحح بسبب هذا التلفيق. وطريقتهم في ذلك هي تضمين الاستبيان مجموعة من الإجابات التي تبدو جذابة للشخص الذي يريد أن يحصل على درجة «جيدة» (أو تبدو دالة على التوافق السَّيىء جيداً للشخص الذي يحاول الحصول على درجة «سيئة»)، ولكن الخبرة قد بينت أنه نادراً ما يختار هذه الإجابات الأشخاص الذين يعطون الإجابات الحقيقية. ومن ثم فكلما زاد اختيار الفرد لمثل هذه الإجابات ازداد الاحتمال بأنه يقوم بتلفيق إجاباته.

ولكن هذا لا يحل كل المشكلة الخاصة بالحصول على تقرير حقيقي. فالفرد قد يكون في كثير من الحالات غير قادر على وصف دوافعه وخصائصه الانفعالية بأمانة حتى ولو رغب في ذلك وقصد إليه. وكما نبهنا فرويد، فإن جزءاً كبيراً من دافعيتنا يعتبر لا شعورياً، وإن قوة كبت الرغبات اللاشعورية تقوم بتشويه وجهة نظر الفرد عن شخصيته. فالرجل ذو الحاجات الاتكالية القوية كثيراً ما يرى نفسه عدوانياً، وفردي النزعة واثقاً من نفسه. والمرأة ذات الكراهية الزائدة في علاقاتها مع أسرتها قد ترى نفسها مهذبة ومنكرة لذاتها. وعلى ذلك، فإن أية درجة قد لا تعني ما يبدو أنها تعنيه.

وفضلا عن ذلك، فإن الدرجات قد تحدد إلى حد ما بأنواع متعددة من الاتجاهات العقلية للاستجابة. وأحد هذه الاتجاهات العقلية للاستجابة هو

الرغبة في القبول الاجتماعي، وهو الميل إلى اختيار الاستجابات التي تعكس ما يعتبر طريقة «صحيحة» للسلوك والشعور في مجتمعنا. واتجاه عقلي آخر للاستجابة هو المسايرة، وهو الميل إلى الموافقة على ما يقوله شخص آخر، أي تفضيل الإجابة «بنعم» على الإجابة «بلا». ومع أن الدراسات والبحوث قد أوضحت أن الاتجاهات العقلية للاستجابة لا تؤثر على وجه عام على الدرجات في اختبارات الشخصية بالقدر الذي يخشى منه، إلا أنه يجب دائمًا أن نكون على حذر منها في الحالات الفردية.

وفضلًا عن ذلك، فهناك تعقيد آخر في محاولة تقدير خصائص الشخصية عن طريق سؤال الناس عن أنفسهم قد أصبح أكثر وضوحاً. ففي هذه الأيام حيث تقدمت وسائل حفظ المعلومات واسترجاعها، قد أصبح كثير من الناس قلقين حول ما يمكن أن يلحق بهم في المستقبل نتيجة إجاباتهم على أسئلة «حساسة» مثل الجنس، أو الدين، أو السياسة. ولهذا فمن الواجب على مؤلفي اختبارات الشخصية ومستخدميها أن يأخذوا في اعتبارهم بجدية حق الأفراد في السرية. فمن المفروض ألا يطلب من الفرد أن يستجيب لاختبار من اختبارات الشخصية إلا لسبب علمي أو شخصي واضح، كما يجب أن تتضمن خطة الاختبار الأصلية توفير السرية أو عدم ذكر اسم صاحب البيانات. والطرق التي تحقق ذلك متوفرة ومتاحة، واستخدامها يعتبر مسؤولية خلقية لمن يقوم بتطبيق اختبارات الشخصية.

وهناك صعوبة أخرى يقابلها الإخصائيون النفسيون في تصميم اختبارات الشخصية وهي نقص المحكات الواضحة لتقدير صدق الاختبار. فكما سبق أن رأينا، فإن هناك طريقة مقننة لتقدير صدق اختبارات القدرات وهي حساب معامل الارتباط بين درجات الاختبار ومؤشرات غير اختبارية للسمة التي يقيسها الاختبار. وقد بدأ ذلك باختبارات الذكاء، حيث نجد أن بينيه قد أخذها كقضية مسلم بصحتها أن السمة التي كان يحاول قياسها هي نفس السمة التي يلاحظها المعلمون في الطلاب النابهين والمتوسطين والأغبياء. وعلى ذلك، فقد أمكنه أن يتأكد من أن اختباراته تقيس هذه الخاصية بوساطة إيجاد معامل الارتباط بين درجات الاختبار وتقديرات المعلمين كما يعبر عنها بالدرجات المدرسية والانتقال إلى صف أعلى، وما شابه ذلك.

أما في حالة معظم خصائص الشخصية التي قد نريد قياسها، فإن هناك

صعوبة في الحصول على مقاييس لمحكات من الحياة الواقعية. إننا نستطيع، بالطبع، ان نسأل الأفراد أن يقدروا تلاميذهم، أو موظفيهم، أو زملاءهم، أو جيرانهم بالنسبة للتفاؤل، أو الاعتماد على الغير، أو التوافق العام. ولكن مثل هذه التقديرات ليست مرضية كمقاييس للشخصية من وجوه كثيرة. أولاً، غالباً ما يختلف القائمون بالتقدير بشكل واضح في تقييمهم للفرد الواحد، وهم بذلك، ربما يعكسون اتجاهاتهم الخاصة بقدر ما يعكسون خصائص الفرد الذي يقومون بتقديره. وثانياً، تتأثر التقديرات بما يسمى بتأثير «الهالة» - بمعنى أن الشخص الذي يستقطب على وجه عام اتجاهات موجبة يميل إلى أن يحصل على تقديرات عالية في كل شيء؛ والشخص الذي يكون على وجه عام غير محبوب يحصل على تقديرات منخفضة في كل شيء. وبالإضافة إلى ذلك، فإنه من الصعب أن تقنع القائمين بالتقدير بأن يوزعوا تقديراتهم على مدى واسع من الدرجات؛ فإذا كان مقياس التقدير يتراوح من ١ إلى ٩، فإنهم يترددون في إعطاء أي فرد ١ أو ٩، وربما يعطون الجميع تقديرات ٥، أو ٦، أو ٧. ولو كانت التقديرات أكثر إقناعاً كطريقة لقياس الشخصية، فربما لم تكن هناك حاجة ملحة لبناء اختبارات الشخصية. وعلى أية حال، فما زالت التقديرات تستخدم على نطاق واسع، سواء لتقييم الشخصية في المواقف العملية أو لاكتشاف العلاقات بين اختبارات الشخصية والمقاييس الأخرى. فالإنسان لا يستغني عن أداة قديمة إلا بعد إيجاد أداة أفضل. ولكن عندما نفسر التقديرات يجب أن نأخذ في اعتبارنا دائماً نواحي القصور فيها. فهي ليست حقيقة محكات مُرْضية لاختبارات الشخصية.

ومن الطرق المستخدمة لقياس صدق اختبار الشخصية طريقة مقارنة مجموعات معينة من الأفراد حددها المجتمع بناء على سمات معينة. وقد كانت هذه الطريقة مفيدة على وجه خاص في البحوث في مجال الميول العصابية والذهانية، إذ أنه من المتوقع أن من هم على درجة كبيرة من سوء التوافق سيلجأون في وقت ما للعلاج الطبي النفسي. وبافتراض أن هناك تدريجاً متصلاً للتوافق، وأن المرضى النفسيين كمجموعة يمثلون إحدى نهايتي هذا التدريج المتصل، فإذن يصبح إثبات حصولهم على درجات متطرفة على اختبار للتوافق دليلاً على صدق هذا الاختبار. (هذا تبسيط زائد. فهناك أنواع كثيرة من سوء التوافق تمثلها مجموعات ذات أمراض نفسية مختلفة، وعلى ذلك فيجب أن نفترض

مجموعة من السمات لا سمة واحدة. ولكن نفس المدأ يمكن تطبيقه على الموقف الأكثر تعقيداً).

. . لقد كان لوجود مجموعات محددة بدرجة واضحة لتسهيل تحديد درجة صدق اختبارات التوافق ـ عدم التوافق أثر كبير في تنشيط البحوث في هذا المجال، بالإضافة الى حاجة مستشفيات الأمراض العقلية والعيادات إلى هذه الاختبارات. ولكن هذا التركيز في البحوث أدى إلى وجود عدد من اختبارات الشخصية في السمات السلبية أكبر من الاختبارات في السمات الإيجابية. فمن السهل جداً أن نكتشف، على سبيل المثال، أن فرداً ما عنده قلق قوي كامن، وميل للتفكير البارانويي، وكثير من الأعراض النفسجمية (السيكوسوماتية)، وهو أسهل من أن نقدر درجة اعتماده على الغير، وإخلاصه، وقدرته على الزعامة. ولذلك يجب أن نتذكر دائماً هذا التحيز السلبي لمقاييس الشخصية كلما استخدمنا هذه المقاييس.

ولقد صحح إلى حد ما هذا التحيز السلبي باستمرار البحوث. ففي ميدان قياس الميول المهنية استفاد إ. ك. سترونج E. K. Strong من وجود مجموعات مهنية في بناء قائمة سترونج للميول المهنية[1]، وهي اختبار قيم لقياس الدافعية التي تحدد قرارات الإنسان في حياته. وفي برنامج آخر كبير للبحوث درس ديفيد ماكليلاند David McClelland وآخرون دافع الإنجاز، كما صمموا بعض الطرق لقياسه. وسوف نتكلم عن هذين الاختبارين كثيراً فيما بعد. أما الآن فسوف نستخدمهما فقط كمثال لاتجاه في البحوث سوف يؤدي يوماً ما إلى التقدم في مجال قياس الشخصية بحيث يمكننا أن نقدر السمات الإيجابية بنفس درجة الكفاءة التي نقيس بها السمات السلبية.

وباستمرار البحوث حول استبيانات الشخصية خلال هذه السنوات العديدة، حدث تقدم في اتجاهات عديدة. فأولاً، بدلاً من أن يكون الاستبيان عبارة عن خليط من الأعراض المتنوعة، كما كان الشأن في استبيانات «التوافق» السابقة، فقد صممت الآن أدوات يمكن أن تعطينا درجات متعددة على سمات مختلفة. فعلى سبيل المثال، إن استبيان مينسوتا المتعدد الأوجه للشخصية

(١) Strong Vocational Interest Blak.

(MMPI)[1]، وهو استبيان معروف جداً، قد بدأ بتسعة مقاييس يمكن عن طريقها تحديد أنواع مختلفة من الاضطرابات النفسية. وبمضي الوقت أضيفت مقاييس أخرى كثيرة تقيس متغيرات ليست مرضية مثل الانطواء الاجتماعي. وقائمة إدواردز للتفضيل الشخصي[2] تقيس خمسة عشر نوعاً من حاجات الشخصية، مثل الحاجة إلى النظام، والحاجة إلى الانتماء، وما إلى ذلك.

ومن الطرق الأساسية التي تتكون بها مفاتيح التصحيح المتعددة لنفس المجموعة من بنود الاختبار هي طريقة تحليل البنود. وهذا يقتضي تبويب إجابات الأسئلة التي أعطاها فعلاً أفراد الجماعات التي اختيرت خاصة لتمثل سمة الشخصية المراد قياسها. فالإجابات التي يوجد بالنسبة لها فرق ذو دلالة احصائية واضحة بين هذه المجموعة وبين مجموعة تمثل المجتمع العام، هذه الإجابات تكوّن مفتاح التصحيح لسمة الشخصية هذه. ففي حالة استبيان مينسوتا المتعدد الأوجه للشخصية الذي أشرنا إليه سابقاً، مثلاً، نجد أن استجابة ما تعطي درجة على مقياس البارانويا إذا ما كان فقط مرضى البارانويا يعطون هذه الاستجابة أكثر مما يعطيها المرضى العاديون. وهذه تسمى الطريقة التجريبية لبناء اختبار في الشخصية - بمعنى ان الخبرة تلعب دوراً أكبر من النظرية أو الفهم العام في تحديد الاستجابات التي توافق كل درجة.

إن وجود هذه الطريقة التجريبية في بناء مفاتيح التصحيح قد أدى إلى اتجاه آخر هام في قياس الشخصية. فقد أصبح واضحاً أنه في كثير من المواقف، تكون الأداة التي تفصّل وتصمم خاصة لمشروع بحث خاص، أو لموقف خاص من مواقف الاختيار - تكون هذه الأداة عادة أكثر كفاءة من اختبار مقنن صمم لقياس السمات العامة للشخصية. فعلى سبيل المثال، لم ينجح الباحثون في تحديد متغيرات الدافعية المتصلة بالتخلف التحصيلي في المدرسة إلا عندما استخدموا الطريقة المشار إليها سابقاً، حيث تمت المقارنة فعلاً بين استجابات الطلاب المتفوقين والطلاب المتخلفين بالنسبة لكل بند من بنود المقياس.

وهناك طريقة رئيسية أخرى يمكن عن طريقها إيجاد مفاتيح التصحيح المتعددة لمجموعة شاملة من بنود اختبارات الشخصية، وهي طريقة التحليل العاملي. وإجراءات هذه الطريقة تتلخص أولاً في أن نقوم بتطبيق بنود الاختبار

(١) Minnesoto Multiphasic Personality Inventory (MMPI).

(٢) Edward Personal Preference Schedule.

على مجموعة كبيرة تكون على قدر الإمكان ممثلة للمجتمع الأصلي الذي يصمم من أجله الاختبار ونحصل على إجاباتها على بنود الاختبار. وتكون الخطوة التالية هي حساب معاملات الارتباطات بين الإجابات على كل بند والإجابات على كل بند من البنود الأخرى، وبذلك نحصل على جدول ضخم من الارتباطات بين البنود. وباستخدام الطرق التي شرحناها في الفصل السابق، يقوم التحليل العاملي بترجمة هذه الارتباطات إلى عوامل يمكن تسميتها عن طريق فحص نظام تشبعات العامل. وقد كان رايموند ب. كاتل Raymond B. Cattell من زعماء مؤيدي هذه الطريقة في بناء اختبارات شاملة في الشخصية.

وكأمثلة لبعض المنجزات في ميدان قياس الشخصية سوف نستعرض استبيانين واسعيّ الانتشار هما: قائمة سترونج للميول المهنية (SVIB)، واستبيان كاليفورنيا السيكولوجي (CPT)[1].

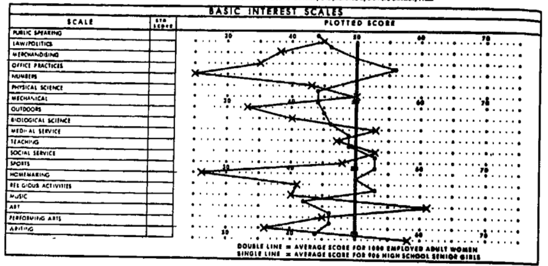
إن قائمة سترونج للميول المهنية أداة تصحح بطريقة تجريبية ذات تاريخ طويل ومشهور. فبعد الحرب العالمية الأولى بقليل لاحظ أ. ل. سترونج وبعض علماء النفس الآخرين حقيقة تثير الانتباه وهي أن المجموعات المهنية المختلفة تختلف فيما بينها اختلافات ثابتة فيما يقولونه عما يحبونه وما لا يحبونه. وبعض هذه الاختلافات كان مما يمكن التنبؤ به بمجرد البديهة. فمن الطبيعي مثلاً، أن يقول عدد أكبر من المهندسين أنهم يحبون الفيزياء إذا ما قورنوا بمجموعة من البائعين. ولكن وجدت كذلك فروق واضحة تخص بعض الأشياء التي لا صلة واضحة لها بالعمل ـ أشياء تتصل بالتسلية، والهوايات، والناس، والكتب، وكثير جداً من نواحي الحياة المختلفة. وتوحي مثل هذه البيانات بأن المهنة قد تمثل طريقة في الحياة، كما تمثل أيضاً طريقة في كسب العيش. وقد رأى سترونج أنه من الممكن قياس هذه الخصائص التي تتصل باختيار المهنة. وفي برنامج منظم للبحث استمر عدة سنوات كثيرة، قام باختبار مجموعات من الرجال من مهن مختلفة، وقارن الاستجابات التي يعطونها على أسئلة الاختبار بالاستجابات التي يعطيها الرجال على وجه عام. فعند بناء مفتاح لمقياس الهندسة المعمارية، مثلاً، طلب من عدة مئات من الذين يزاولون مهنة الهندسة المعمارية أن يجيبوا على الاختبار. ثم قام بتسجيل إجاباتهم على كل بند على حدة ليرى أي البنود

(١) California Psychological Inventory (CPI).

أجابوا عليها بكلمة «أحب»، وأيها (غير مهتم) وأيها «لا أحب». وإن أي إجابة تختلف عن إجابة الناس عموماً ويكون لهذا الاختلاف دلالة إحصائية فإن هذه الإجابة توضع ضمن مفتاح تصحيح مقياس الهندسة المعمارية.

ثم قام سترونج فيما بعد بتصميم صورة خاصة من الاختبار للنساء، وكون مقاييس خاصة بالمهن النسائية بنفس الطريقة. وقام بإعداد معايير لجميع المجموعات المهنية التي شملتها الدراسة، كما قام بدراسة العلاقة بين درجات الميول وكل من العمر الزمني، والقدرات الخاصة، وكثير من الخصائص الإنسانية الأخرى. وتتألق قائمة سترونج للميول المهنية للرجال في صورتها الحالية من ٣٩٩ بنداً أخذت من ميادين مختلفة من الحياة - المهن، المواد الدراسية، نواحي التسلية، أنماط البشر، مواقف العمل، وهكذا. أما قائمة سترونج للميول المهنية للنساء فهي تتألف من بعض نفس البنود، بالإضافة إلى البنود الأخرى المميزة لأعمال النساء في مجتمعنا. وتقوم المؤسسات التي تتولى مهمة تصحيح الاختبارات باستخدام ٥٤ مقياساً مهنياً لقائمة الرجال، و٢٢ مقياساً للميول الأساسية التي تقيس الميول لأنواع خاصة من الأنشطة (مثل الخطابة، والعلوم، والمغامرات)، وبعض المقاييس الإضافية غير المهنية لأشياء مثل دافعية الإنجاز الأكاديمي، والانطواء - الانبساط. أما فيما يختص بقائمة النساء فإن هذه المؤسسات تقوم في العادة باستخدام ٥٨ مقياساً مهنياً، و١٩ مقياساً للميول الأساسية، وأربعة مقاييس إضافية. وقد شرح معنى الدرجات المختلفة باختصار خلف قائمة صحيفة البروفيل التي يتسلمها المفحوص، كما أنه شرح شرحاً اوفى في الدليل التوضيحي. ويبين شكل رقم ١٨ كيف يبدو بروفيل الفرد.

وبناء على البحوث الشاملة التي امتدت أكثر من أربعة قرون فقد عرفنا مجموعة من الأشياء عن الميول المهنية كما تقيسها قائمة سترونج. فأولاً، إن نمط الأشياء التي يحبها الإنسان أو التي لا يحبها ليست أساساً نتيجة الانخراط في مهنة ما، ولكنها موجودة من قبل الالتحاق بالمهنة. فعلى سبيل المثال، نجد أن طلبة جامعة ستانفورد الذين التحقوا بكلية الطب فيما بعد وأصبحوا أطباء، قد - ارا على درجات عالية على مقياس الأطباء في قائمة سترونج حتى أثناء دراستهم في السنوات الجامعية الأولى وقبل التحاقهم بكلية الطب. ويبدو أن معظم الناس يكونون أنماط ميولهم الشخصية قبل أن يتركوا المدرسة الثانوية.



OCCUPATIONAL SCALES

	OCCUPATION	STD. SCORE		OCCUPATION	STD. SCORE
I	MUSIC TEACHER		VII	ARMY-ENLISTED	
	ENTERTAINER			NAVY-ENLISTED	
	MUSICIAN PERFORMER			ARMY-OFFICER	
	MODEL			NAVY-OFFICER	
II	ART TEACHER		VIII	LAWYER	
	ARTIST			ACCOUNTANT	
	INTERIOR DECORATOR			BANKWOMAN	
III	[illegible]SWOMAN			LIFE INS. UNDERWRITER	
	ENGLISH TEACHER			BUYER	
	LANGUAGE TEACHER			BUSINESS ED. TEACHER	
IV	Y.W.C.A. STAFF MEMBER		IX	HOME ECON. TEACHER	
	RECREATION LEADER			DIETITIAN	
	DIRECTOR CHRISTIAN ED		X	PHYSICAL ED TEACHER	
	NUN, TEACHER			OCCUPATIONAL THERAPIST	
	GUIDANCE COUNSELOR			PHYSICAL THERAPIST	
	SOCIAL SCIENCE TEACHER			PUBLIC HEALTH NURSE	
	SOCIAL WORKER			REGISTERED NURSE	
V	SPEECH PATHOLOGIST			LIC. PRACTICAL NURSE	
	PSYCHOLOGIST			RADIOLOGIC TECHNOLOGIST	
	LIBRARIAN			DENTAL ASSISTANT	
	TRANSLATOR		XI	EXECUTIVE HOUSEKEEPER	
VI	PHYSICIAN			ELEMENTARY TEACHER	
	DENTIST			SECRETARY	
	MEDICAL TECHNOLOGIST			SALESWOMAN	
	CHEMIST			TELEPHONE OPERATOR	
	MATHEMATICIAN			INSTRUMENT ASSEMBLER	
	COMPUTER PROGRAMMER			SEWING MACHINE OPERATOR	
	MATH SCIENCE TEACHER			BEAUTICIAN	
	ENGINEER			AIRLINE STEWARDESS	

NON OCCUPATIONAL SCALES: 51 (AACH) 40 (DIV) 45 (FEM) 39 (OIE)

ADMINISTRATIVE INDICES: 380 (TR) 3 (UR) -1 (SE) 36 (LP) 2 (IP) 60 (DP)

شكل ١٨ ـ بروفيل افتراضي على قائمة سترونج للميول المهنية للنساء يبين ميول طالبة في الدراسات العليا في علم النفس. وتبين درجاتها في الميول الأساسية أنها ميالة إلى العلم، والتدريس، والموسيقى، والكتابة. وتوحي المقاييس المهنية أن علم النفس ميدان جيد لها.
وفي البروفيل الخاص بالميول الأساسية (الجزء الأعلى) يعبر الخط المزدوج عن الدرجة المتوسطة ل ١٠٠٠ امرأة عاملة. ويعبر الخط المنفرد عن الدرجة المتوسطة ل ٩٠٦ طالبة بالسنة النهائية بالمدرسة الثانوية.
(Reproduced from the 1969 Supplement to the Manual for the **Strong Vocational Interest Blank,** Stanford University Press).

ثانياً، يلاحظ عند معظم الناس، أن مثل هذه الميول عندما تتكون تصبح ثابتة ودائمة شأنها في ذلك شأن أي خاصية من خصائص الشخصية التي درست. وقد استمر سترونج على اتصال بطلبة جامعة ستانفورد الذين طبق عليهم الاختبار في الثلاثينات، وكان يكتب إليهم من حين إلى آخر ويطلب منهم الإجابة على الاختبار مرة أخرى، حتى يمكن أن يعرف ما إذا كانوا قد تغيروا. وعلى الرغم من أنه وجد بعض التغيرات الطفيفة كما هي العادة، وعلى الرغم من أن بعض الأفراد تغيرت صورة ميولهم تماماً في أوقات مختلفة، إلا أن الغالبية العظمى لم تتغير بدرجة واضحة على مدى ٢٢ سنة.

ويمكن القول أيضاً أن البحوث ذات الفترات الطويلة قد أوضحت الكثير عما يمكن لدرجات قائمة سترونج أن تتنبأ به وما لا يمكن أن تتنبأ به. وباستثناء بعض الحالات فإن هذه الدرجات لا تساعدنا على التنبُّؤ بمدى نجاح الفرد المحتمل في مهنة ما أو في البرنامج التدريبي الذي يؤدي إليها. (ربما لا يجب أن نتوقع ذلك على الإطلاق، فدرجة الفرد توضح إلى أي مدى يشبه هذا الفرد مجموعة مهنية مكونة كلها من رجال ناجحين بدرجة تكفي لبقائهم في هذه المهنة. ولكن درجات النجاح لم تؤخذ في الاعتبار). ولكن ما يمكن للدرجات أن تتنبأ به هو مدى احتمال بقاء الفرد في مهنة معينة أو احتمال انتقاله من هذه المهنة إلى مهنة أخرى. وفضلاً عن ذلك، فإنه على الرغم من أن بعض الدراسات توضح أن معاملات الارتباط بين درجات الميول وبين رضا الفرد عن مهنته كما يقرر هو ذلك بنفسه ليست عالية جداً، فإن هناك أدلة تشير إلى أن الذين يعملون في مهن قد حصلوا فيما يتعلق بها على درجات عالية في قائمة سترونج يكونون على وجه عام أكثر رضاء عن مهنهم من الذين لم تتفق درجات ميولهم مع الاتجاه المهني الخاص الذي اتجهوا إليه.

وهناك استبيان آخر للشخصية استخدم على نطاق واسع في العديد من البحوث والمواقف العملية، وهو استبيان كاليفورنيا السيكولوجي (CPI) (تشير المراجع التي جمعها مؤلف الاستبيان هاريسون ج. جف Harrison G. Gough إلى أكثر من ٦٠٠ دراسة نشرت حتى أوائل عام ١٩٦٩). فمن أجل إعداد اختبار يفيد في كثير من المواقف ـ مثل المدارس، والعيادات النفسية، ومعاهد التأهيل (الإصلاح)، ومستشفيات الأمراض العقلية ـ بدأ جَفْ بإعداد البنود التي تمثّل «المفاهيم الشعبية» أي تلك الآراء التي تتعلق بالشخصية التي اعتاد الناس على

استخدامها كلما أرادوا تقدير شخصية بعضهم البعض أو حاولوا التنبؤ بما يمكن أن يفعله الآخرون، ولم يستخدم المفاهيم التي تقوم على النظريات المتخصصة في الشخصية او تصنيفات الطب النفسي. وباستخدام أساليب تحليل البنود لمقارنة فئات خاصة من الناس بالناس العاديين ـ مثل مقارنة الأحداث المنحرفين بالأحداث الأسوياء من البنين والبنات من نفس السن ـ تمكن جف من تكوين استبيان يقيس ثماني عشرة سمة مختلفة:

أولاً ـ مقاييس التوازن، والتسلط، وتأكيد الذات، وكفاءة العلاقات الشخصية:

١ ـ السيطرة.

٢ ـ المكانة.

٣ ـ المقدرة الاجتماعية.

٤ ـ الحضور الاجتماعي.

٥ ـ تقبل الذات.

٦ ـ الإحساس بالوجود في حالة جيدة.

ثانياً ـ مقاييس التطبيع الاجتماعي، والنضوج، والمسؤولية، وبناء القيم لدى الفرد:

٧ ـ المسؤولية.

٨ ـ التطبيع الاجتماعي.

٩ ـ ضبط النفس.

١٠ ـ التحمل.

١١ ـ الانطباع الجيد.

١٢ ـ المشاركة الاجتماعية.

ثالثاً ـ مقاييس قدرات الإنجاز والكفاءة العقلية:

١٣ ـ الإنجاز من خلال المسايرة.

١٤ ـ الكفاءة من خلال الاستقلالية.

١٥ ـ الكفاءة العقلية.

رابعاً ـ مقاييس الصيغ العقلية وصيغ الميول:

١٦ ـ العقلية السيكولوجية.

١٧ ـ المرونة.

١٨ ـ الأنوثة.

ويبين الشكل رقم ١٩ كيف يبدو بروفيل درجات أحد الأفراد.



Name ________ Age ________ Date Tested ________

Other Information ________

Notes:

FEMALE NORMS

Do Cs Sy Sp Sa Wb Re So Sc To Gi Cm Ac Ai Ie Py Fx Fe

Standard Scores

Female Norms

شكل ١٩ ـ بروفيل لمصيفة جوية على استبيان كاليفورنيا السيكولوجي يوضح السمات التي يبدو أنها هامة في هذه المهنة ـ التوازن، والتحمل، والاستقلالية، والمرونة.

(Reproduced with-permission from the Manaul for the **California Psychological Inventory** by Harrison G. Gough, P. 15 Copyright 1959 by Consulting Psychologists Press Inc Palto Alto, Calif).

إن ثبت المراجع الشامل الذي أشرنا إليه سابقاً يمدنا بدليل على أن استبيان كاليفورنيا السيكولوجي يمكن أن يستخدم فعلاً في الأعراض العملية التي قام المؤلف بتصميمه من أجلها ـ فيمكن، مثلاً، أن يتنبأ بأي من مجموعة من الطلبة النابهين لا يحتمل أن يفكر في الالتحاق بالجامعة إلا إذا بُذِلَ مجهود خاص لجذب انتباهه إلى ذلك، أو بأي من شباب إحدى المناطق الفقيرة من المحتمل جداً أن ينحرفوا إذا لم توفر لهم مساعدة أو رعاية خاصة. وشيء آخر جانبي طريف بالنسبة لهذا الاختبار، وهو أنه ترجم إلى عدة لغات أخرى أوروبية، وأسيوية، وقد وجد أن بعض المقاييس ذات قدرة تنبئية في إيطاليا، واليابان، والهند، مثلما وجد في الولايات المتحدة بالنسبة لهذه المقاييس.

الأساليب الإسقاطية

سوف نستعرض الآن الاتجاه الرئيسي الآخر في اختبار الشخصية، وهو اتجاه الأساليب الإسقاطية. وفي هذه الأساليب يقوم الفاحص، بدلاً من أن يسأل المفحوص مجموعة من الأسئلة، بعرض مثير غامض عليه ويطلب منه تفسيره أو الاستجابة له. وبقع الحبر وصور من المواقف الإنسانية كانت الأكثر استخداماً (انظر الشكل رقم ٢٠).



شكل ٢٠ ـ صورة مشابهة لتلك التي تستخدم في اختبارات الشخصية الإسقاطية.

وطالما أن بقعة الحبر ليست في الحقيقة صورة لشيء ما، فإن التفسير الذي يعطيه الفرد لا بد أن يكون نابعاً من داخله، وبالتالي فإنه يعبر بعض الشيء عن الطريقة التي بها يدرك وينظم عالمه الخاص. وبالمثل، طالما أن الصورة المبينة في شكل ٢٠ لا توضح ماذا يفعل هذان الشخصان أو ماذا يقول كل منهما للآخر، فإن أية قصة يرويها الفرد عنها لا بد أن تنبعث من صميم أفكاره ومشاعره. وهذا يشير إلى السبب في تسمية هذا النوع من الاختبارات. فيقال إن المفحوص **يسقط** على هذه الصورة اتجاهاته الانفعالية الخاصة وآراءه في الحياة. وقد اقترح البعض أنه قد يكون من الأفضل تسمية هذه الأساليب بالطرق **المعبرة عن الذات**، ولكن اسم **الإسقاطية** قديم، ولذلك استمر شائع الاستخدام.

إن أول هذه الاختبارات والذي استخدم على نطاق واسع، والذي ربما ما زال أكثرها استخداماً حتى الآن، هو اختبار رورشاخ. وهذا الاختبار عبارة عن مجموعة من عشر بقع من الحبر من مختلف الأشكال. والبعض منها ابيض وأسود فقط، والبعض الآخر ملون أيضاً. ويسأل المفحوص أن يحكي ماذا يرى في كل واحدة منها. وبعد أن يقول المفحوص كل ما يريد أن يقوله، يقوم الفاحص بتوجيه ما يريد من أسئلة ليستوضح ما رآه المفحوص في البقعة وأي ناحية فيها حددت إدراكاته. ومنذ عام ١٩٢١ حينما اقترح هرمان رورشاخ Herman Rorschach، الطبيب النفسي السويسري، هذه الطريقة، اقترحت عدة طرق كثيرة ومختلفة لتصحيح هذا الاختبار، كما كتب عدد كبير من الكتب والمقالات في المجلات العلمية حول هذا الاختبار. ولقد اهتمت طرق التصحيح المختلفة بتحليل بناء أو أسلوب الاستجابات، ثم مضمونها، كل منها على حدة. أما **البناء** فيتعلق بمثل هذه الأسئلة: إلى أي مدى كان المفحوص طلقاً في الاستجابة لهذا المثير؟ أو بمعنى آخر، هل يعطي المفحوص عدداً كبيراً من الاستجابات لكل صورة أو يعطي استجابات قليلة؟ هل هو يستجيب عادة لكل شكل بأكمله أم يستجيب لجزئياته؟ إلى أي مدى يعتمد في استجاباته على الشكل، أو اللون، أو الظلال؟ أما **المضمون** فإنه يتصل بمدى رؤية المفحوص للبقع كأشكال بشرية، أو حيوانات، أو رسوم تشريحية، أو خرائط، أو سحب، أو أي أشياء أخرى عيانية. كما أن هناك اهتماماً أيضاً بمدى ميل المفحوص إلى إعطاء استجابة شائعة أو استجابات مبتكرة.

ويجب أن يكون واضحاً حتى من هذا الوصف المختصر للنواحي الكثيرة التي يجب أن تحلل في استجابات الفرد لبقع الحبر، أن عملية التفسير إنما هي عملية على درجة كبيرة من التعقيد. وذلك لأنه ليست هناك طريقة يمكن بها إثبات صدق اختبار كهذا بطريقة حاسمة. وباستمرار البحث على مدى سنوات كثيرة ظهر أن كثيراً من الاستنتاجات التي اعتاد أن يقوم بها الإخصائيون الإكلينيكيون على أساس هذا الاختبار لم تكن سليمة. فلم يثبت أن هذا الاختبار قادر على كشف أسرار الشخصية بالطريقة التي كان يظنها أكثر المتحمسين له. ولكن على الرغم من وجود بعض النقائص، فإنه لا يزال أداة ذات قيمة للإخصائيين العاملين في ميدان علم النفس الإكلينيكي. فإذا استخدم هذا الاختبار بالإضافة إلى الطرق الأخرى لتقدير الشخصية مثل المقابلة والمعلومات عن خلفية المفحوص، فإنه سوف يمدنا بمعلومات حول موضوعات يمكن أن يكون من المفيد استكشافها في دراسة حالة فردية ما. وفي عام ١٩٦١ ظهر اختبار بقع الحبر لهولتزمان Holtzman ليصحح بعض النقائص من اختبار رورشاخ ولكن مع الاحتفاظ بكثير من مزاياه.

وهناك اختبار إسقاطي آخر يستخدم كذلك على نطاق واسع هو اختبار تفهم الموضوع[١]. ويتألف هذا الاختبار من مجموعة من البطاقات عليها صور للناس. وقد ترك الموقف في كل صورة وكذلك مشاعر الأفراد فيها من غير تحديد - عمداً - حتى يمكن أن يعبر المفحوص عما يراه بناء على اتجاهاته وطرق إدراكه للعالم. إذ أن المطلوب من المفحوص أن يؤلف قصة لكل صورة بحيث تتضمن تفسيراً للأسباب التي أدت إلى الموقف المبين في الصورة، وكذلك ما يفكر فيه الأشخاص وما يشعرون به، ثم النتيجة المحتملة لهذا الموقف. ويبين الشكل رقم ٢٠ صورة مماثلة للصورة المستخدمة في اختبار تفهم الموضوع.

ومثلما حدث بالنسبة لاختبار رورشاخ، فقد ظهرت طرق عديدة لتحليل وتفسير قصص المفحوصين في اختبار تفهم الموضوع. وعلى وجه عام، فإن الفاحص يبحث أولاً ليرى أي شخصية أصبحت بطل القصة، حيث أنه من

---

(١) Thematic Apperception Test (TAT). ظهرت نسخة عربية لهذا الاختبار حيث أعيد رسم معظم صوره لكي تلائم البيئة العربية. قام بإعداد هذه النسخة العربية للاختبار محمد عثمان نجاتي وأنور حمدي. القاهرة: دار النهضة االعربية، ١٩٦٦.

المحتمل أن يكون هو الشخصية التي يتوحد معها المفحوص عندما يروي هذه القصة. ثم يقوم الفاحص بعد ذلك بدراسة تفاصيل كل قصة بعناية ـ ماذا قال المفحوص، وكيف قال ما قال ـ حيث يكوّن بعض الفروض عن حاجات الفرد، ومشاعره، واتجاهاته، وحيث يلاحظ كذلك مدى تكرار ظهور نفس الاتجاهات في القصص المختلفة. ويستخدم كثير من الاخصائيين الإكلينيكيين هذه المعلومات كأساس لتكوين وصف لغوي للمفحوص أكثر منه لوضع درجة أو مجموعة من الدرجات. أما الباحثون، فإنهم يستخدمونها لإعطاء تقديرات شاملة لسمات الشخصية التي يهتمون بدراستها.

وقد اشتغلت إمكانات البحث في هذه الطريقة العامة في عدة ميادين مختلفة، حيث صممت مجموعات من الصور خاصة بالأطفال، وأخرى بالزنوج، وثالثة لأفراد من ثقافات مختلفة، كما قننت طرق التصحيح بالنسبة لبعض الخصائص النوعية. ومن الأمثلة الجيدة على استخدام أسلوب رواية القصص في البحوث ما قام به ماكليلاند ومعاونوه في دراساتهم لدافعية الإنجاز والتي أشرنا اليها سابقاً (انظر صفحة ١١١). وقد قام الباحثون بعرض مجموعة من الصور التي تم اختيارها بعناية على المفحوصين من أجل اثارة تأليف القصص عن الإنجاز، ثم قاموا بعد ذلك بتحليل قصص كل مفحوص بناء على نظام مقنن. فقد أخذوا في اعتبارهم الرغبات التي تبديها الشخصية الرئيسية، والأنشطة التي تقوم بها، والعوائق التي تقابلها، بالإضافة إلى كثير من المواضيع الأخرى. وعلى هذا، فقد أمكنهم الحصول على درجة لكل مفحوص بالنسبة للحاجة إلى الإنجاز. وفي برنامج للبحث استمر مدة طويلة درسوا الطرق التي من خلالها ارتبطت هذه الدرجات بخلفية المفحوصين وإنجازاتهم التالية.

وفي إمكاننا أن نستعرض الكثير من الأنواع الخاصة للاختبارات الإسقاطية، ولكن أي مناقشة مسهبة سوف تأخذنا بعيداً عن اهتماماتنا الحالية. وهناك شيء واحد يمكن أن نقوله عن الاختبارات الإسقاطية جميعها: وهو أن محاولات إيجاد معامل صدق لها لم تكتمل بعد. ولذلك، فإن هذه الاختبارات لا يمكن استخدامها الآن في اتخاذ قرارات عن الآخرين في المواقف العملية كما يجب أن يفعل المعلمون، أو الإخصائيون الاجتماعيون، أو مديرو التوظيف. ولكن إذا استخدمها إخصائي نفسي مقتدر ذو خلفية معرفية جيدة عن بناء ووظيفة الشخصية، فإنها يمكن أن تساعد في كثير من الحالات في اتخاذ مثل هذه

القرارات. ومن هنا كانت هذه الخلفية ضرورية في اكتساب مهارة استخدام الاختبارات الإسقاطية، وليس هناك بديل عنها. لذلك فإن دراسة مقرر واحد في الأساليب الإسقاطية لا تكفي، بالطبع، لاكتساب هذه المهارة.

وربما كانت القيمة الكبرى للأساليب الإسقاطية تكمن فيما تقدمه من إسهامات في بحوث الشخصية، حيث أنها تساعد الإخصائيين النفسيين على اكتشاف بعض المؤشرات على أشياء لا يستطيع الناس ان يتكلموا عنها مباشرة - مثل دوافعهم الداخلية، وافتراضاتهم، والطرق التي ينظرون بها إلى العالم.

من القياس الى التقدير

إننا نسمع قليلاً هذه الأيام عن قياس الشخصية، ولكننا نسمع كثيراً عن تقدير الشخصية، فإلى أي نوع من التغير في تفكيرنا يرجع هذا التغير في التسمية؟ إن ذلك، أولاً، وقبل كل الشيء، اعتراف بأن الشخصية نظام معقد بالفطرة ومتعدد الجوانب. ولا يبدو أنه من المحتمل أننا سنصبح يوماً ما قادرين على أن نميز الشخصية بدرجة رقمية أو بمجموعة من الدرجات. لذلك، فإن مصطلح التقدير يعطي تصوراً أوسع وأشمل لشخصية الفرد أكثر مما يعطي مصطلح القياس.

ثم إنه بالإضافة إلى الاختبارات المقننة، فإن أكثر الطرق المستخدمة شيوعاً في عمليات التقدير هذه هي الملاحظات، والمقابلات، ومقاييس التقدير. إن نقطة الضعف الرئيسية في هذه الطرق انما ينشأ عن أن شخصية كل من الفاحص والمفحوص تتدخل في عملية التقويم. ولكن إذا علمنا أن أي وصف لشخص ما باستخدام طريقة واحدة فقط قد يكون متحيزاً أو غير دقيق، فإنه يمكن أن نستخدم مجموعة من هذه الطرق للحصول على بيانات تؤدي إلى تقديرات سليمة للشخصية.

إن البحوث في ميدان تقدير الشخصية قد أدت إلى إيجاد أساليب يمكن أن يدمج فيها كل من طريقتي الملاحظة والقياس. فالبنود التي تعتمد على تاريخ الحياة مثل تلك التي تحتويها طلبات الاستخدام، مثلاً، يمكن أن تصحح باستخدام طريقة مقارنة المجموعات، على نحو ما اتبع اثناء عملية تكوين مفاتيح التصحيح التجريبية لاستبيانات الشخصية. فعند مقارنة مندوبي شركات التأمين الناجحين، مثلاً، بمجموعة أخرى من غير الناجحين في هذا العمل، فإنه من الممكن

التعرف على بعض البنود الهامة في طلبات **الاستخدام**، التي لها قيمة تنبئية بالنسبة لهذه المهنة. وهناك طريقة أخرى لدمج كلتا الطريقتين الجديدة والقديمة معاً، وذلك بأن يقوم الباحث بالحصول على تقديرات لمفحوصيه يقوم بوضعها أشخاص آخرون، مع مراعاة تقنين **الموقف** الذي تتم فيه ملاحظتهم، فإذا جمع المفحوصون معاً في جماعة للمناقشة أو للقيام بعمل يثير التحدي فإن الملاحظات التي تتم في هذه **المواقف تكون** أكثر دقة من تلك التي تتم في المواقف المختلفة في الحياة اليومية.

وقد بحثت طرق أخرى عديدة لتقدير الشخصية، ولكن الأمثلة السابقة تعتبر كافية لتوضيح أن هناك وسائل أخرى بجانب الاختبارات لتقدير الفرد - كيف يفكر في العادة، وكيف يشعر، وكيف يستجيب لظروف الحياة. فليس من المهم ان نستطيع قياس الشخصية، ولكن المهم ان نستطيع تقديرها.

بعض الاعتبارات العامة

إن علماءالنفس قد اضطروا، نتيجة لتقدم البحوث وتراكم نتائجها إلى الاعتراف بأن السلوك الذي نعتبره **مظهراً** لسمات الشخصية لا يمكن التنبؤ به بدرجة عالية. ويبدو أن السبب في ذلك هو أن الفرد يكتسب أثناء نموه **مجموعة** من الأدوار التي يستطيع أداءها، اي **مجموعة (تشكيلة)** من «شخصيات» مختلفة على نحو ما. فالطريقة التي يتصرف بها الفرد تكون في كثير من الحالات، أقرب اتصالاً بالموقف الذي يجد نفسه فيه، منها إلى حاجاته الداخلية ودوافعه. فنحن ندرك أن الفرد (أ) يظهر نوعاً من «الشخصية» عندما يتحدث إلى زوجته، ويظهر شخصية مختلفة تماماً عندما يتكلم مع رئيسه في العمل. ونحن نعرف كذلك أن أدق ما نستطيع القيام به من تقديرات الشخصية لن تساعدنا في التنبؤ، ولو بدرجة معقولة من الدقة، كيف سيتصرف مريض ما في مستشفى الأمراض العقلية عندما يغادر المستشفى، إلا إذا عرفنا جيداً طبيعة الموقف الذي سيكون فيه. إن أساليب تقدير الشخصية تساعدنا على فهم الناس وفهم أنفسنا، ويجب ألا نتوقع منها أكثر من ذلك.

الفصل السابع

تطبيقات الاختبارات والمقاييس

الاختبارات والقرارات بشأن الافراد

الاختبارات هي عبارة عن أدوات صممت لتستخدم في اتخاذ القرارات البشرية. وفي مجتمعنا المعقد المتعدد الجوانب تتخذ كل يوم الاف القرارات التي تنطوي على بعض التقييم للخصائص النفسية للأفراد. وبعض هذه القرارات يتخذها الناس لتنظيم حياتهم الخاصة، كأن يقرر فرد ما أن يلتحق بالدراسات العليا بدلاً من أن يلتحق بعمل في جيولوجيا البترول، أو كأن تقرر فتاة ما الالتحاق بوحدة عسكرية نسائية. وبعض هذه القرارات يتخذها بعض الناس في شأن البعض الآخر، مثل عندما يقرر أحد مديري الأعمال أي واحدة من ثلاث متقدمات سيختار لتكون سكرتيرة، او عندما يضع طبيب في عيادة للأمراض العقلية ومساعدوه خطة علاج مريض جديد. وفي أي من هذه المواقف، إذا توفر وجود اختبار جيد في يدي شخص يفهمها، فإنه سوف يكون ذا فائدة كبيرة.

لقد أكدنا في الفصول السابقة أن عملية الإفادة من المعلومات المستمدة من الاختبار إنما هي عملية معقده، ذلك يرجع إلى أنه ليس هناك اختبار يعتبر مقياساً صادقاً صدقاً كاملاً لخاصة نفسية معينة، وأن الدرجة ليست بالضرورة مقياساً دقيقاً لأي شيء - بمعنى آخر، إن صدق وثبات أي اختبار ليسا كاملين. ومن أجل ذلك نجد أن برامج الاختبارات والإرشاد النفسي كثيراً ما تسير جنباً إلى جنب. فمناقشة نتائج الاختبار مع إخصائي يفهم جيداً ما يمكن استنتاجه وما لا يمكن استنتاجه من هذه النتائج سوف يؤدي إلى قرارات سليمة لا يمكن الوصول إليها عن طريق مجرد الاعتماد على الدرجات فقط. فعلى سبيل المثال، عندما يجابه فرد ما مشكلة المفاضلة بين أن يقبل وظيفة جيولوجي في شركة من شركات البترول أو أن يلتحق بالدراسات العليا للحصول على درجة الدكتوراه، فربما أراد هذا الفرد أن يقارن قدرته العقلية العامة بالقدرة العقلية العامة لطلبة الدراسات العليا الآخرين عن طريق اختبار مثل اختبار الخريجين. ولكن كيف سيستخدم المعلومات الخاصة بأنه حصل على درجة ٤٥٠؟ فإذا كانت جداول المعايير لهذا الاختبار تدل على أن هذه الدرجة أقل قليلاً من المستوى المتوسط لطلبة الدراسات العليا في الجيولوجيا، فإنه لا يمكن أن يكون متأكداً أنه سوف يحصل على درجات أقل من المتوسط. فنتائج الطلبة الذين سبقوه تبين أن بعض الطلبة الذين حصلوا على درجات في الاختبار مثل درجته كان متوسط تقديراتهم في الدراسات العليا «ممتاز» (أ)، بينما البعض الآخر كان متوسط تقديراتهم «جيد منخفض» (- ج). فإلى أي فئة سوف ينتمي هذا الفرد؟ إن أي خطأ في التقدير نحو أي من هذ . الاتجاهين عند اتخاذه القرار سوف يكلفه الكثير. فإذا قرر أن يلتحق بالدراسات العليا، ثم بعد مضي سنة أو سنتين يكتشف أنه لم يستطع أن يكون مؤهلاً للحصول على درجة الدكتوراه، فإنه سوف يشعر أنه أضاع جزءاً هاماً من حياته. وفضلاً عن ذلك، فقد لا تتوفر الفرصة حينذاك للحصول على وظيفة جيدة مماثلة لتلك التي تفرضها عليه شركة البترول الآن. ومن جهة أخرى، إذا قرر عدم الالتحاق بالدراسات العليا بناء على درجة دون المتوسط التي حصل عليها، فربما يكون قد أضاع على نفسه طوال بقية حياته فرصة الوصول الى أعلى المستويات في مهنته، تلك المستويات التي قد يصل اليها على الأقل بعض

الأشخاص الذين يحصلون على درجات تماثل درجته. فإذا أراد هذا الفرد أن يتخذ قراراً حكيًما في مثل هذا الموقف، فلا بد أن يأخذ في اعتباره عوامل كثيرة. وإن درجته في الاختبار عامل واحد فقط من هذه العوامل.

إن الاختبار الذي يلعب دوراً في عملية اتخاذ قرارات حيوية يجب أن يتم اختياره بدقة. فمع أن الاختبار يجب أن يكون ثابتاً بدرجة كافية لكي يوضح لنا بدقة الوضع النفسي للفرد فيما يتعلق بالخصائص التي يقيسها، فإن الاعتبار المهم في الحقيقة هو الصدق التنبئي. ويجب أن نعرف أي نوع من التقدم في بعض الميادين الخاصة، كما ظهر من نتائج البحوث، ما يمكن أن نتوقعه من الأفراد ذوي الدرجات المنخفضة والدرجات المتوسطة والدرجات العالية. إن أداة الاختبار التي تستخدم في اتخاذ القرارات لا يتم بناؤها كلها دفعة واحدة مثل بناء منزل أو آلة. ولكنها تنمو على مدى حقبة من السنوات، كما تزداد الفائدة منها بازدياد الخبرة بها. ويمكننا أن نرتب جميع الاختبارات المختلفة التي تكلمنا عنها في مقياس يبين مدى نمو كل منها في الوقت الحالي. إن الاختبار الجديد، مهما كانت فكرته بارعة، ومهما كانت الحاجة الاجتماعية إليه كبيرة، فإنه سوف يكون في أدنى مقياس «النمو» هذا. وإن اختباراً ما مثل اختبار ستانفورد بينيه الذي ظل يستخدم بصورة دائمة منذ عام ١٩١٦ لا بد أن يأتي في أعلى المقياس، إذ أن الإخصائي المدرب الذي له معرفة بالمعلومات الكثيرة التي تراكمت حول هذا الاختبار، يعرف ماذا يقيس هذا الاختبار وأي أنواع من القرارات يمكن أن تتخذ بناء عليه.

ومع ذلك، فهناك دائًما فرصة لظهور اختبارات جديدة في ميدان الإرشاد النفسي، حيث يمكن أن تكون في بعض الحالات مصدرا مفيداً للأفكار سواء للمرشد النفسي أو لطالب الإرشاد للتحقق من صحة بعض المعلومات المستمدة من غير مواقف الاختبارات. ومع ذلك، فإنه من المهم على وجه خاص، في مثل هذه الحالات، أن تكون لدى المرشد النفسي المعرفة الكافية لكي يستطيع أن يميز بوضوح أنواع نتائج الاختبارات التي لا يمكن الاعتماد عليها بمفردها.

وهناك نقطة عملية أخرى يجب أن تؤخذ في الاعتبار عندما تتدخل نتائج الاختبارت في القرارات الشخصية، وهي أنه من المحتمل أن تكون الدرجات المنخفضة ذات دلالة حتى ولو لم يكن الأمر كذلك بالنسبة للدرجات العالية ـ والعكس صحيح. فإذا عدنا إلى حالة الفرد الذي يريد الالتحاق بالدراسات العليا أو يقبل

وظيفة جيولوجي في شركة من شركات البترول، فإنه عندما يجد أن درجة ٥٠٠ هي الحد الأدنى للقبول في الدراسات العليا التي يرغب في أن يتلقى فيها برنامجاً ممتازاً في الجيولوجيا، يصبح من الواضح له أن الدرجة التي حصل عليها وهي ٤٥٠ تعني أنه غير مؤهل للالتحاق بهذه الدراسات، بصرف النظر عن مزاياه الأخرى الكثيرة. ولكن، إذا كان قد حصل على درجة ٦٥٠ بدلاً من ٤٥٠، فلن تكون له ميزة تجعله أفضل من متقدم آخر حصل على ٥٥٠ درجة، حيث أنه بعد مستوى معين من القدرة العقلية العامة يصبح النجاح في الدراسة العليا معتمداً على خصائص أخرى عقلية وخلقية غير الاستعداد الأكاديمي. فمثل هذه المعلومات تمثل جزءاً هاماً من الخلفية التي يحتاجها الفرد لكي يفسر نتائج الاختبار بذكاء. إن معرفة معاملي الصدق والثبات وحدهما لا تكفي.

أما مناقشة بعض الموضوعات الأخرى الأقل وضوحاً والتي يجب أن تؤخذ في الاعتبار عند استخدام الاختبارات في تقييم الذات فهي تتصل بموضوع جدلي كان يثار من وقت لآخر طوال السنوات الماضية. فهل يجب أن نخبر الأطفال بنسب ذكائهم؟ فمن جهة، يقول البعض إن من حقهم وحق آبائهم أن يعرفوا حتى يتمكنوا من التخطيط واتخاذ القرارات المناسبة وخاصة فيما يتعلق باستمرار الدراسة. ومن جهة أخرى، يقول البعض الأخر إن نسبة الذكاء إنما هو رقم، وبدون الخلفية التي أشرنا إليها يصبح من المحتمل جداً إساءة تفسيره، وبالتالي يؤدي إلى خطط غير سليمة. وربما كانت أفضل سياسة يحسن على المدرسة ان تتبعها هي أن تجعل نتائج الاختبارات في متناول التلاميذ وأولياء الأمور على أن يتم توفير المعلومات التفسيرية اللازمة مع هذه النتائج. ومما يجدر أن نشير إليه هو انه من المشكوك فيه أن تستطيع أية مدرسة الاستفادة من برامج الاختبارات ما لم يكن ضمن الهيئة الفنية فيها من يفهم الاختبارات وحدودها. وغالباً ما يقوم المرشدون النفسيون بهذه الوظيفة.

القرارات التي تخص الآخرين

إن أنواع القرارات التي تتخذ بالنسبة للآخرين والتي من أجلها شاع استخدام الاختبارات فيما مضى إنما هي قرارات اختيار الأفراد. فعندما يكون المطلوب من الانسان أن يختار فرداً أو أكثر من بين مجموعة من المتقدمين، فإن

ما يسهم به الاختبار يمكن أن يحلل بطريقة سهلة ومباشرة. فإذا كان الاختيار للدراسة الجامعية أو لعمل، او للانضمام إلى مجموعة علاجية، فإن هدف من يتخذ القرار هو أن يختار الفرد المناسب للمكان المناسب له حسب تصوره. وهو لا يتوقع أن يكون مصيباً في كل مرة يتخذ فيها القرار، ولكنه يريد أن يكون عدد الناجحين الذين يختارهم أكثر من عدد الفاشلين. وتستطيع الاختبارات في بعض الحالات تحسين مستوى الأفراد المختارين، حتى وإن كانت معاملات ثباتها وصدقها التنبئي ليست عالية بدرجة كبيرة.

جدول ٦ ــ كيف يساهم اختبار مهارة الكتابة بالاختزال من اختيار مرشحين جيدين

العدد في كل مجموعة من مجموعات الدرجات موزعون حسب التقديرات في القدرة على الاختزال				درجات اختبار مهارة الكتابة بالاختزال	النسبة المئوية في كل مجموعة من مجموعات التقديرات موزعة حسب مجموعات درجات الاختبار			
دون المتوسط	متوسط	فوق المتوسط	ممتاز		دون المتوسط	متوسط	فوق المتوسط	ممتاز
	٤	٦	٧	١٨ - ١٩		١٧	٤٠	٦٤
	٢	٢	٤	١٦ - ١٧		٩	١٣	٣٦
	١٠	٥		١٤ - ١٥		٤٤	٣٣	
	٤	١		١٢ - ١٣		١٧	٧	
٢	٢	١		١٠ - ١١	٦٧	٩	٧	
١	١			٨ - ٩	٣٣	٤		
٣	٢٣	١٥	١١		١٠٠	١٠٠	١٠٠	١٠٠

يوضح جدول التوقع هذا العدد والنسبة المئوية للكاتبات بالاختزال الحاصلات على تقديرات مختلفة على القدرة على الاختزال موزعات حسب درجاتهن في اختبار مهارة الكتابة بالاختزال. (العدد ٥٢، المتوسط ١٥,٤، الانحراف الميعاري ٢,٩، ر=٠,٦١؛ والدرجة هي عبارة عن المتوسط لمجموعات من خمسة حروف). ويبين هذا الجدول أنه لو وظفت فقط الحاصلات على ١٤ درجة على الأقل، لما حصلت أي منهنّ على تقدير «دون المتوسط»، ولحصل كثير منهنّ على تقدير «فوق المتوسط» أو «ممتاز».

ولنأخذ، على سبيل المثال، جدول التوقع المبين في الجدول رقم ٦. فإذا كان الشخص المسؤول عن التوظيف سيختار فقط المتقدمين الحاصلين على ١٤

درجة فما فوق في الاختبار، فإنه يمكنه أن يتوقع أن ينجح جميع المتقدمين ولا يفشل منهم أحد. وعندما يقل عدد العمال المتقدمين عن المطلوب فإنه قد يضطر الى تشغيل العمال الحاصلين على أقل من ١٢ درجة حتى يستطيع الحصول على عدد كاف من العمال لكي يضمن استمرار العمل. وفي هذه الحالة، وعلى الرغم من أنه يعلم مقدماً أن عدد الفاشلين سوف يكون أكثر، إلا أن هذا العدد كان من الممكن أن يكون أكثر لو أنه لم يستخدم الاختبار. إن الاختبار، بالنسبة لمديري التوظيف، إنما هو أساساً أداة تساعد على ازدياد عدد العاملين الناجحين، وتقلل عدد العاملين الفاشلين. وهو لا يهتم كثيراً بمعنى الدرجات او نسق هذه الدرجات. وهو يعلم كذلك أن هناك بعض الأخطاء في الاختيار تحدث مع استخدام الاختبار - فبعض المتقدمين الحاصلين على درجات عالية لا يعملون جيداً - ولكنه لا يقلق بالنسبة لحالات الفشل هذه طالما ان عددها قليل. كما إن هناك خطأ آخر محتملاً ولكنه لا يهتم به على الإطلاق - فقد لا يوظف عدداً من الأشخاص بسبب درجاتهم المنخفضة في الاختبار، وكان من الممكن أن يكونوا عمالاً ممتازين.

وعلى أية حال، فقد ازداد عدد الناس الذين أخذوا يقلقـون على مـا للطريقة المقننة التي تستخدمها الاختبارات في قرارات الاخـتيـار من آثار بعيدة المدى على المجتمع ككل. فمدير شؤون الأفراد قد لا يهتم بأمر الأفراد الذين حصلوا على درجات منخفضة والذين كان من المحتمل أن يكونوا ناجحيّن في أعمالهم لو أنهم وظفوا، ولكن المجتمع يهتم بأمرهم. فكثير من الحالات التي تعاني من البطالة المزمنة هي لأشخاص استبعدوا مرة بعد أخرى لأنهم لم يحصلوا أبداً على درجات قريبة من الدرجة المتوسطة (قمة توزيع الدرجات) في أي اختبار، سواء في المدرسة أو في مكاتب التوظيف. لقد أصبح من الواضح أن ما نستخدمه من طرق للاختيار لا تؤدي إلى وضع كل إنسان في عمل ما.

ومن الممكن أن تكون أهم نتيجة لاهتمامنا المتزايد بالآثار غير المباشرة للاختبارات على المجتمع هي تحوّل في المفاهيم من نظام الاختيار إلى نـظام التصنيف،وذلك عندما نفكر في اتخاذ قرارات بشأن توظيف الأفراد. فبدلاً من أن نسأل: هل يقبل هذا الفرد أو يرفض؟، يكون السؤال: في أي عمل يكون هذا الفرد أكثر كفاءة؟ وبناء على ذلك تختفي كل المشاكل التي أشرنا إليها فيما سبق.

وبعض الشركات التقدمية تقوم الآن بإجراء التجارب حول خطط للتوظيف تعمل على توظيف كل من يرغب في العمل، ثم بعد ذلك تستخدم كل مصادر الإدارة المهنية للأفراد للوصول إلى قرار حول أي عمل أو برنامج تدريبي يكون أنسب لهذا الفرد أو ذاك. وغالباً ما تتبع وكالات التوظيف نفس هذا الأسلوب، وخاصة الوكالات التي تعنى عناية خاصة بتوظيف الأحداث، أو غير ذوي الخبرة، أو المعوقين.

ويمكن أيضاً أن نعتبر القرارات التشخيصية التي نتخذها بناء على استخدام الاختبارات في العيادات النفسية هي قرارات بشأن التصنيف أكثر منها قرارات بشأن الاختيار، إذ أن السؤال الأساسي هو: «ما هي المشكلة هنا؟». فعندما يدخل أحد المرضى العقليين المستشفى، فإن الخطوة الأولى في علاجه هي محاولة اكتشاف ما يعاني منه. فإذا كانت الأعراض العامة التي يشكو منها تشير إلى احتمال وجود حالة فصام أو ذهان عضوي ناتج عن مرض في الدماغ، فإن الإخصائي النفسي يلجأ إلى استخدام اختبارات قد بيَّنت البحوث السابقة أنها تميز بين مرضى الفصام ومرضى الذهان العضوي. وحيث أن مثل هذه القرارات التشخيصية يقوم بها فريق من الخبراء وليس الإخصائي النفس بمفرده، فإن الأدلة التي نحصل عليها من اختبارات لا بد أن يساندها ما نحصل عليه من أدلة صادرة من أنواع أخرى من الفحص. وفي هذه الحالة يجري تحليل للرسم الكهربائي للمخ (تسجيل النشاط الكهربائي في المخ)، ويفحص تقرير الطبيب عن الأمراض التي أصيب بها المريض والتي يحتمل أن تكون قد سببت تلف الدماغ. ويستخدم الاختبارات أيضاً بهذه الصفة التشخيصية الإخصائيون في القراءة الذين يحاولون اكتشاف مواطن الضعف عند القارىء البطيء وذلك من أجل تعليمه ما لم يتعلمه في الفصل المدرسي، كما يستخدمها أيضاً المتخصصون في علاج أمراض الكلام الذين يرون أن خصائص الشخصية ذات علاقة بحالة التهتهة، كما يستخدمها أيضاً الكثير من العاملين في الميدان الإكلينيكي والتربوي.

وتختلف مثل هذه الاستخدامات الإكلينيكية عن مواقف التوظيف التي أشرنا إليها سابقاً من حيث أن نتائج الاختبارات تمدنا فقط بمؤشرات أو فروض نقوم دائمًا بالتحقق من صحتها بمقارنتها بأنواع أخرى من المعلومات المتوفرة عن الشخص موضع الدراسة. وفي الحقيقة إن استخدام الاختبارات هنا في

المواقف الإكلينيكية قد يكون له قيمة أكثر إذا ما قورن باستخدامها في مواقف التوظيف، حتى ولو لم تكن على درجة عالية من الصدق والثبات بالنسبة للغرض الذي يستخدمها الفاحص من أجله، حيث أن المهم هو براعة الفاحص في التفكير في الطرق التي تمكنه من تتبع المؤشرات الجديدة التي تتصل بمتاعب المريض. فقد يبتدع بعض طرق الاختبار التي لم تسبق دراستها على الإطلاق دراسة منظمة. فإذا أمدته هذه الطرق بفروض يستطيع أن يتحقق من صحتها بمقارنتها ببعض الحقائق المستمدة من سلوك المريض، أو أعراض مرضه، أو تاريخه، فإن هذه الطرق تكون قد أدت الغرض منها. ولكن يجب أن نقول إنه إذا نجحت طريقة القصة غير المقننة كاختبار في حالة معينة (أو في حتى عشر حالات أو مائة حالة) فإن هذا لا يعتبر دليلاً على أنها أداة مفيدة في اتخاذ القرار في كل الحالات المشابهة. وفي الحقيقة، إن كثيراً من الأدوات المستخدمة في الاختبارات التشخيصية لا تتمشى مع المعايير التي أشرنا إليها في الفصل الثالث. وهذا ينطبق أيضاً على نسبة كبيرة من الاختبارات الإسقاطية المستخدمة في العيادات الإكلينيكية. ولذلك، يجب أن نكون على حذر من أن نستنتج أن هذه الأدوات سوف تكون ذات قيمة إذا كانت القرارات سوف تتخذ كلية أو اساساً بناء على نتائج الاختبارات فقط. فإنه من الضروري أن يتم التحقق من الفروض التي بنيت على نتائج الاختبارات التي لا تكون دقيقة دقة كافية، وذلك بمقارنة هذه النتائج بمصادر أخرى للمعلومات.

الاختبارات والقرارات: بعض الاعتبارات العامة

من الواضح أن المعايير التي نطبقها عندما نختار اختباراً ما يجب أن تحددها الفائدة التي نبتغي تحقيقها من نتائج هذا الاختبار. فنحن لا نستطيع أبداً أن نقول: «هذا هو أفضل اختبار للذكاء، لذلك يمكن استخدامه لكل غرض»، أو «هذا اختبار جيد للنضج الانفعالي». ولكننا نسأل دائمًا: «مَنْ سوف يستخدم نتائج هذا الاختبار، وفي أي الأغراض»؟.

ويجب أنَ نجرب الاختبار المقترح في الموقف المعين الذي نتوقع أن نستخدمه فيه كلما سنحت الفرصة لذلك. وبعد أن نعرف أداء المفحوصين في المجموعة التجريبية، يمكننا أن نكون جداول توقع مثل تلك الجداول المبينة في الجدول رقم ٦، أو أن نقوم بوضع معادلات انحدار من أجل الأغراض التنبئية

على أساس هذه المجموعة. فعلى سبيل المثال، إذا كانت شركة من شركات التأمين قد وظفت في الماضي عدداً قليلاً من فئة معينة من الموظفات الكتابيات، ثم ترغب الآن في زيادة عددهن، فإنه من المستحسن إجراء دراسة أولية يُطبَّق فيها على جميع المتقدمات اختبار القدرة الكتابية الذي كانت الشركة تستخدمه، ولكنها تقوم بتوظيفهن بدون الاعتماد على نتائج الاختبار. وبعد ستة شهور، مثلاً، يمكن أن يقارن أداء هؤلاء الموظفات بدرجاتهن على الاختبار. إن ذلك سوف يبين كم كان عدد جميع الفاشلات، وكم من هؤلاء كان من الممكن أن يُستبعَدْنَ لو حددت درجة معينة للنجاح تبين عدم صلاحية بعض المتقدمات منذ البداية (وسوف يبين ذلك أيضاً كم من هؤلاء اللاتي اتضح أنهن موظفات أكفاء وكان من الممكن أن يُستبْعَدن ولا يعطون الفرصة لإظهار أنهن في الواقع أكفاء). وكثيراً ما قامت برامج كبيرة للاختبار بدون مثل هذه التجارب الأولية وأدت الى قرارات غير سليمة في شأن الناس.

وكثيراً ما يكون مفيداً أن نخطط لدراسات أولية من أجل مقارنة صلاحية عدة اختبارات بدلاً من تقييم اختبار واحد فقط. كما أنه من المستحسن، إذا كان هدفنا هو التصنيف وليس الاختيار، أن نخطط لدراسات أولية تشمل المتقدمين لأعمال مختلفة متعددة بدلاً من المتقدمين لعمل واحد فقط. ففي المثال الذي ذكرناه في الفقرة السابقة كان من الممكن استخدام خمسة اختبارات قصيرة من أجل اختيار هؤلاء الموظفات الكتابيات لوظائف مثل الكتابة بالاختزال، ومسك الدفاتر، وتثقيب البطاقات، وكتبة الملفات.

والأشخاص المسؤولون عن اتخاذ القرار عما إذا كانت الاختبارات ستستخدم أم لا تستخدم في موقف معين من مواقف اختيار الأفراد، يجب عليهم في كثير من الحالات أن يقارنوا تكاليف برنامج الاختبار بتكاليف التدريب. فإذا كان العمل يتطلب مهارة يمكن أن يتعلمها أي فرد في أيام قليلة، فإن استخدام الاختبارات لاختيار الأفراد الملائمين يصبح أمراً لا يستحق العناء الذي يُبذَل فيه. أما إذا كان الأمر يحتاج إلى فترات طويلة من التدريب وإلى أدوات مكلفة من أجل إكساب الفرد المهارة المطلوبة، كما في حالة الطيارين، مثلاً، فإن برنامج الاختبار، حتى ولا كان مكلفاً جداً يصبح أمراً له ما يبرره إذا كان سوف يؤدي في الواقع إلى اختيار المرشحين الذين سيصبحون في نهاية الأمر ناجحين.

إن اصدار الأحكام بالنسبة للآخرين عملية معقدة. فالاختبارات ما هي إلا أدوات يستخدمها الذين تكون وظيفتهم إصدار مثل هذه الأحكام. فالاختبارات لا تصنع القرارات لنا، ولا تستطيع أن تؤدي الدور بمفردها. فإذا تذكرنا دائمًا هذه الحقيقة، فإننا سوف نجني ثمار فوائد الاختبارات النفسية دون أن نقع في أخطائها.

الاختبارات كأدوات للبحث

لقد بينا في الفصل الأول أن القياس أساسي للبحث النفسي. والقياس، بطبيعة الحال، مصطلح عام يشمل الكثير من الطرق بالإضافة الى تلك الطرق التي نسميها عادة الاختبارات. دعنا الآن ننظر باختصار إلى هذا السؤال. كيف يمكن استخدام الاختبارات من أجل تقدم المعرفة في علم النفس؟.

الفروق الفردية

إن أقرب ميادين البحث إلى القياس العقلي هو ميدان علم النفس الفارق، أي دراسة الفروق بين الأفراد. فبالإضافة إلى تصميم الأدوات التي تؤدي أغراضاً مفيدة في المدارس، والعيادات، ومكاتب التوظيف، قام الإخصائيون النفسيون بتقديم عدد من النظريات، كما أجروا البحوث للتحقق من صحتها ـ نذكر منها، مثلاً، النظريات حول طبيعة الذكاء، وبناء القدرات العقلية، وأصول الفروق بين الجنسين، والعلاقة بين الخصائص العقلية والجسمية.

ولعل من أهم ما تركه هؤلاء الباحثون الأوائل للباحثين الحاليين الذين يدرسون الفروق الفردية هو إيضاح التمييز بين المسائل المفيدة وغير المفيدة. ففي البداية كانت الأسئلة واسعة وشاملة، مثل هل الفروق الفردية في الذكاء ناتجة عن الوراثة أو البيئة؟ هل الرجال أكثر ذكاء من النساء؟ هل يختلف الذكاء باختلاف الأجناس؟ هل نسبة الذكاء ثابتة؟ هل يختلف الناس في خصائصهم النفسية أكثر من اختلافهم في خصائصهم الجسمية؟ هل تميل القدرة العقلية للجنس البشري إلى الانحدار؟ هل تزيد الفروق بين الأفراد أو تقل نتيجة للتدريب؟.

وبمضي الوقت اصبح واضحاً أن الاختبارات العقلية لن تجيب على هذا

النوع من الأسئلة، وذلك لأسباب متعددة. فبعض المشكلات مثل مقارنة الاختلاف في ذكاء الإنسان بالاختلاف في الطول (أي أطوال الأفراد) وجدت أنها غير قابلة للحل وذلك بسبب خواص درجات المقياس، بمجرد أن فهمت هذه الخواص. فالطول يقاس على مقياس النسبة (انظر الفصل الأول) بحيث يمكن أن نقول أن أطول إنسان. يبلغ طوله ثلاثة أمثال أقصر إنسان. ولكن الذكاء لا يقاس على مقياس النسبة بحيث أن قسمة نسبة ذكاء على نسبة ذكاء أخرى تعطي رقمًا عديم المعنى. ولذلك، فإننا لا نستطيع أن نقارن اختلاف الناس في سماتهم العقلية باختلافهم في السمات الجسمية عن طريق استخدام هذا النوع من الاختبارات العقلية المتوفرة لدينا.

وبالنسبة للأسئلة الأخرى، مثل مقارنة الجنسين، أو الأجناس، أو الطبقات الاجتماعية فإن المشكلة في هذه الحالة هي مشكلة محتوى الاختبارات التي نستخدمها لقياس الخصائص العقلية. فالاختبارات تمكننا من إجراء مقارنات صحيحة بين الأفراد في داخل الجماعة الاجتماعية الواحدة لأنه من الممكن أن نفترض أن خبرة هؤلاء الأفراد الذين نطبق عليهم الاختبار متشابهة إلى درجة تكفي لجعل درجات الاختبار تعبر أساساً عن الخاصة العقلية أو الفكرية للأفراد أكثر مما تعبر عن بعض الخصائص غير المألوفة في خلفياتهم. فعندما نقارن متوسطات مجموعات نعرف أنهم يختلفون من حيث خبرتهم بالمشكلات والمسائل المماثلة لتلك التي يتضمنها الاختبار - مثل السود والبيض (في أمريكا) - فإننا سوف لا نعرف معنى النتائج التي نحصل عليها. فعلى الرغم من أن البيض يسجلون في العادة درجات أعلى على اختبارات الذكاء، إلا أننا لا يمكننا أن نستنتج أن الجنس الأبيض أكثر ذكاءً، كما أنه لا يمكننا أيضاً - بناء على نتائج الاختبارات العقلية - أن نقول إن الأجناس متساوية في الذكاء. فنحن لا نستطيع أن نستنتج شيئاً على الإطلاق. وأفضل ما يمكن عمله هو ترك هذه المشكلة.

وفي وقت ما، كان يُظن أن الاختبارات الخالية من العوامل الحضارية أو الاختبارات المتوازنة حضارياً سوف تعيد، هذا الموضوع إلى مسرح الدراسة مرة أخرى، غير أن تقدم البحوث قد قضى على هذه الظنون. فإذا أردنا أن نقدم أسئلة الاختبار في شكل صور لا كلمات، فإن هذا لن يحل المشكلة، لأنه قد أصبح معروفاً الآن أن النظر إلى الصورة إنما هو مهارة مكتسبة، وأن الأطفال في بعض الثقافات لا يكتسبون هذه المهارة. وإذا فكرنا في أن نحول الأسئلة إلى

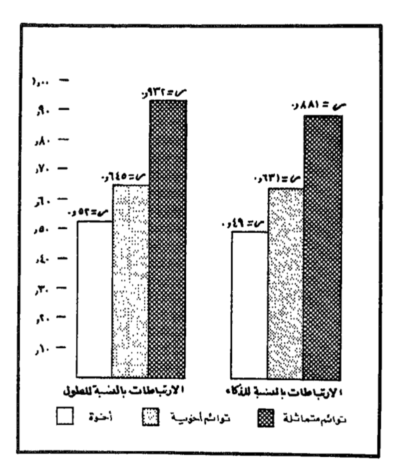
أشكال هندسية بدلاً من الأشياء المألوفة الاستخدام، فإن هذا أيضاً لن يحل المشكلة لأن إدراك الأشكال إنما هو أيضاً مهارة مكتسبة. فالناس الذين يعيشون في بيوت مستديرة الشكل ويقضون النهار في الحقول أو الغابات لا يتعلمون تمييز الزوايا والمستطيلات.

ويبدو الآن أنه من الأحسن إعادة النظر في أسئلة البحوث بحيث تصاغ في صورة أكثر ضيقاً وتحديداً، بدلاً من أن نحاول بناء بعض أساليب الاختبارات من أجل الإجابة على أسئلة واسعة. فإذا أخذنا على سبيل المثال السؤال الذي اهتم به المختصون في علم النفس الفارق أكثر من أي سؤال آخر، وهو الأثر النسبي لكل من الوراثة والبيئة على الفروق الفردية في الذكاء، فإنه يمكننا أن نرى كيف أن برنامجاً للبحوث للإجابة على أسئلة أكثر تحديداً قد أدى إلى مجموعة من المعارف التي يوثق بها على الرغم من أنها ليست كاملة. فقد عرفنا أن هناك فروقاً في القدرات العقلية ترجع إلى الوراثة، حتى ولو لم نحصل على مقاييس خالصة (أو نقية) لهذه الفروق.

إن أكثر النتائج دلالة قد جاءت من دراسة التوائم. فالتوائم المتماثلة أو **أحادية اللاقحة** مادة مثالية لبحث موضوع الوراثة، وذلك لأن هؤلاء التوائم لهم مجموعات متماثلة من المورّثات (الجينات). وفي مثل هذه الحالة ينمو الطفلان التوأم من بويضة واحدة مخصبة قد انقسمت إلى جزءين في وقت حدوث الانقسام الأول للخلية. ولذلك، فإن أية اختلافات في السلوك أو القدرة أو الشخصية بين مثل هذه التوائم يجب أن يرجع إلى بعض تأثيرات البيئة، ولا يمكن أن تكون وراثية.

وفي دراسة التوائم هناك نوعان من تصميمات البحوث يكثر استخدامها. أولهما هو مقارنة مدى التشابه بين أزواج من التوائم المتماثلة بمدى التشابه بين أزواج أخرى من الأطفال. والخطوة الأولى هي أن نكتشف مدى التشابه عموماً بين أزواج من التوائم المتماثلة في خاصة ما كالذكاء مثلاً. والخطوة التالية هي اكتشاف مدى التشابه بين أزواج التوائم الأخوية، أو **ثنائية اللاقحة** (وهم التوائم الذين ينمو كل زوج منهم من بويضتين مستقلتين مخصبتين، وبالتالي فهما ليسا أكثر تشابهاً من حيث نظام المورثات من أي أخين عاديين). وكثيراً ما تضاف خطوة ثالثة ـ وهي اكتشاف التشابه بين الإخوة غير التوائم. وحيث أن التشابه بين أطفال الأسرة الواحدة يمكن أن يكون مصدره الوراثة أو البيئة أو كليهما

معاً، فإن مقارنة أنواع مختلفة من أزواج الإخوة سوف تساعدنا على استخلاص بعض هذه المحددات والوصول إلى بعض الاستنتاجات الأولية عما إذا كانت سمة ما لها أساس وراثي أم لا. فإذا كان لها أساس وراثي، فإننا نتوقع أن يكون هناك شبه أكثر في هذه السمة بين التوائم المتماثلة حيث أن عوامل كل من الوراثة والبيئة تؤثر عليهم بطريقة متماثلة. وفي حالة الإخوة من أعمار مختلفة يكون الشبه بينهم أقل ما يمكن، وذلك لأن تأثيرات البيئة الأسرية على نموهم لا تكون متماثلة بسبب مرور الزمن بين طفل وآخر. أما التوائم الأخوية فيجب أن يكونوا بين هؤلاء وهؤلاء، لأنهم من حيث عوامل الوراثة لا يختلفون عن الإخوة العاديين، ولكنهم يتعرضون لتأثير البيئة الأسرية في نفس الوقت. وفيما يتعلق بالذكاء، فهذا هو في الواقع ما وجدناه تماماً، كما هو مبين في الشكل رقم ٢١.

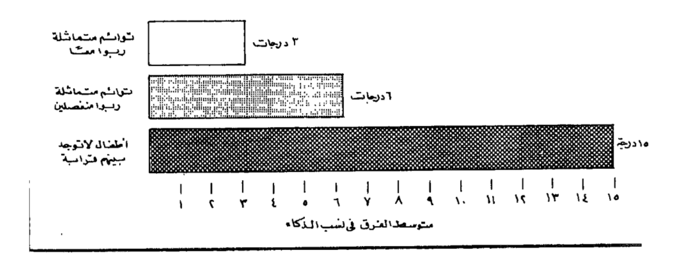


شكل ٢١ ـ معاملات الارتباط بين أزواج الأطفال في نفس الأسرة بالنسبة للطول والذكاء.

(Twin correlations from R. S. Woodworth. **Heredity and Environment.** Soc. Sci. Res. Counicl, New York, 1941. Sibling correlations from H. S. Conrad and H E. Jones. **Yearbook Nat. Soc. Stud. Educ.** 39 (1940), II: 97- 141: and K. Pearson and A. **Lee.** **Biometrika** 2 (1903): (357- 462).

وحيث أن هـذا النظام من الارتبـاطات يشبـه كثيراً ذلـك النظام من الارتباطات في حالة الطول وبعض الخصائص الجسمية الأخرى التي نعرف أن الوراثة هي التي تحددها أساساً، فإنه يبدو أنه من المعقول أن نستنتج أن الفروق في الذكاء أيضاً إنما ترجع إلى الوراثة. وقد وجد نفس النظام في حالة بعض القدرات الخاصة مثل المهارات الميكانيكية والحركية، وفي حالة بعض انحرافات الشخصية مثل الفصام.

أما التصميم الثاني في دراسة التوائم فإنه يتضمن القيام بمقارنات التشابه بين التوائم المتماثلة في حالة تنشئتهم معاً بالتشابه بينهم في حالة تنشئتهم في بيئات مختلفة. ويبين الشكل رقم ٢٢ النتائج. فإذا كانت **الوراثة هي كل شيء** في تحديد ذكاء الفرد أو خصائصه النفسية الأخرى، فإن تنشئة التوائم المتماثلة في نفس البيت والمجتمعَ، أم لا، تصبح أمراً غير ذي أهمية. أما إذا كانت **البيئة هي كل شيء** فإن التشابه بين التوائم المتماثلة الذين نُشِّئُوا في بيئات مختلفة لن يكون أكثر من التشابه بين مجموعة أزواج من الأطفال تؤخذ بطريقة عشوائية. والسبب الأساسي الذي نعرفه الآن في أن كلا هذين الاتجاهين المتطرفين (الوراثة والبيئة) غير صحيح فيما يتعلق بالـذكاء وبعض الخصـائص العقلية الأخـرى هو أن البحوث تشير إلى حقيقة ما بين هذين الطرفين. وكما نرى في الشكل رقم ٢٢، إن متوسط الفرق بين التوائم المتماثلة في حالة التنشئة في بيئات مختلفة أكبر من



شكل ٢٢ ـ متوسط الفرق في نسب الذكاء لأزواج من الأطفال من درجات مختلفة من القرابة.
(From R. S Woodworth **Heredity and Environment.** Soc. Sci Res. Council , New York, 1941).

الفرق بين التوائم المتماثلة في حالة التنشئة في بيئة واحدة، ولكن هذا الفرق أقل كثيراً من الفرق بين الأطفال الذين لا توجد بينهم قرابة.

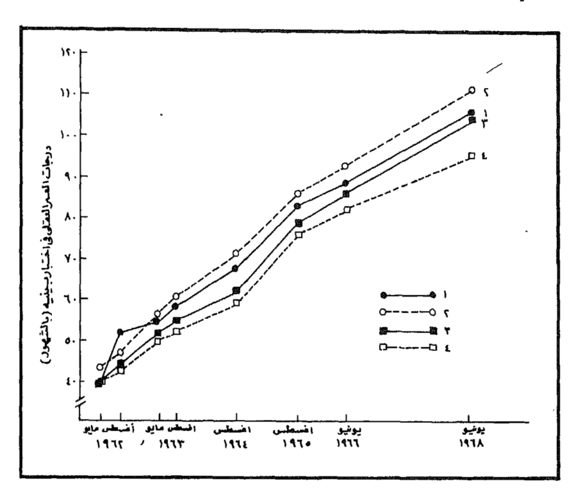
لقد كانت هذه الدراسات عن التوائم المتماثلة في حالة تنشئتهم في بيوت ومجتمعات مختلفة هي مصدر معلوماتنا عن أي مكونات البيئة له الأثر الأكبر على الذكاء. ففي دراسة جادة اتبعت منهج دراسة الحالة قام نيومان H.H. Newman بدراسة ١٩ زوجاً من التوائم المتماثلة الذين تَربَّوا منفصلين. وقد قام الباحث بتقدير بيئاتهم بالنسبة لما تقدمه لهم من عوامل الرعاية التربوية والاجتماعية. وقد وجد أنه في جميع الحالات التي يتفوق فيها توأم على آخر من الناحية العقلية، كان هناك اختلاف واضح في العوامل التربوية. أما عوامل الرعاية الاجتماعية فيبدو أنه لم يكن لها تأثير كبير.

وهناك بعض المسائل الأخرى في موضوع الأثر النسبي للوراثة والبيئة على مختلف الخصائص النفسية قد درست بوسائل عديدة أخرى: مثل دراسات تاريخ الأسر، والبحوث التي تجرى على الأطفال في بيوت الرعاية، ودراسات تربية الحيوانات. وتشير جميع هذه البحوث إلى استنتاج عام وهو أن الكثير من القدرات وسمات الشخصية أو معظمها له أصل وراثي، ولكن نمو هذه القدرات والسمات يتأثر أيضاً بالمؤثرات البيئية.

ويهتم الباحثون في الوقت الحالي بالبحث عن أنواع معينة من المؤثرات البيئية. ففي خلال الستينات ـ على سبيل المثال ـ كان الاهتمام موجهاً إلى التربية المبكرة، أي التعلم الذي يكتسبه الأطفال في مرحلة ما قبل المدرسة والذي عليه يعتمد تعلمهم فيما بعد. ويبين الشكل رقم ٢٣ مثالاً للنتائج التي وصلت إليها سوزان جراي Susan W. Gray وروبرت كلاوس Rupert A. Klaus ، والتي بينت أن معدل النمو العقلي للأطفال الذين يمرون بخبرات تربوية مخططة بعناية في مرحلة ما قبل المدرسة أعلى بشكل واضح من معدل النمو العقلي للأطفال الذين لا يمنحون هذه الخبرات التربوية. أما إلى أي مدى تستمر هذه الزيادة في معدل النمو في السنوات اللاحقة، فهو موضوع لبحوث أخرى خاصة لا زالت البيانات تجمع عنه بوساطة هؤلاء الباحثين وغيرهم في أماكن مختلفة.

إن البحث في الفروق الفردية قد أخذ يتحول بصورة متزايدة إلى بحث في النمو الإنساني. فالمعرفة التي سوف تكون ذات فائدة عظمى في التخطيط لتقدم المجتمع لا بد أن تتصل بأنواع معينة من المؤثرات التي يجب أن تطبق في

مراحل معينة من النمو وفي مواقف محددة. ونحن الآن نحرز بعض التقدم في جمع هذه المعرفة. فالاختبارات العقلية تعمل كمؤشرات لكيفية تقدم نمو الأفراد، وبالتالي فإنها تلعب دوراً هاماً في مثل هذه البحوث.



شكل ٢٣ ـ منحنيات للنمو العقلي خلال فترة ٦ سنوات لمجموعتين من الأطفال في مرحلة ما قبل المدرسة (٢،١) ولمجموعتين ضابطتين (٤,٣)

(From Rupert A. Klaus and Susan W. Gray, unpublished manuscrpt, by permission).

البحوث التجريبية

بعد عدة سنوات كثيرة من بداية حركة القياس النفسي كوّن الإخصائيون النفسيون الذين يقومون بدراسات تعتمد على الملاحظة ومعاملات الارتباط فرعاً مستقلاً في علم النفس العلمي، بينما كوّن الآخرون الذين يقومون بالتجارب في المختبرات فرعاً آخر. وكان التفاعل بين هذين الفرعين قليلاً جداً. فالإخصائيون النفسيون الذين يقومون بالدراسات الارتباطية يستخدمون

الاختبارات، في حين أن الإخصائيين النفسيين التجريبيين يستخدمون عادةً أنواعاً أخرى من المقاييس. والتمييز الأساسي بين هذين الأسلوبين في البحث هو أن الإخصائي النفسي التجريبي يتناول المتغيرات المستقلة. فهو يقدم المثيرات التي يستجيب لها الفرد ثم يقيس التأثير الناتج عن التغير في هذه المثيرات.

وطالما أن الباحثين في المختبرات كانوا يهتمون بالنواحي الأقل تعقيداً في التعلم، والدافعية، والأداء، فقد كان يكفي استخدام مقاييس بسيطة للسلوك ـ مثل تسجيل عدد الأخطاء التي يقوم بها الفأر في محاولة عبوره إحدى المتاهات، أو عدد الأجزاء من المليثانية التي يستغرقها الفرد ليضغط على مفتاح كهربائي استجابة لإشارة معينة. ولكن مع تزايد تطبيق المنهج التجريبي على موضوعات التفكير الإنساني، والشخصية، والسلوك الاجتماعي، أصبحت الحاجة واضحة إلى مقاييس لهذه الخصائص الإنسانية المعقدة، ولذلك اتجه التجريبيون إلى استخدام اختبارات الذكاء، والقدرات الخاصة، والتحصيل، والشخصية. ثم أخذ يتزايد شيئاً فشيئاً اجتياز الإخصائيين النفسيين لذلك الخط الوهمي الذي كان يفصل من قبل بين الباحثين في المختبر ومستخدمي الاختبارات العقلية.

وتستخدم الاختبارات غالباً في التجارب النفسية كمتغيرات تابعة. فقد يهتم أحد الباحثين، على سبيل المثال، بدراسة أثر العقاقير على الأداء في اختبارات الذكاء، أو بدراسة آثار الضغط (أو الشدة) على الثقة بالنفس كما تقاس بالاختبارات. وقد تستخدم الاختبارات أيضاً كمتغيرات ضابطة. فقد يريد أحد الباحثين، مثلاً، أن يدرس أثر الأنواع المختلفة من التعليمات على طريقة حل المشكلة، ومن ثم يبدأ في تقسيم المفحوصين إلى فئة عليا، وأخرى متوسطة، وثالثة دون المتوسط بناء على اختبار من اختبارات الذكاء.

والاختبارات التي تستخدم في قياس المتغيرات في التجارب النفسية أو التربوية يجب أن تقيم بطريقة تختلف عن طريقة تقييم الاختبارات التي تصمم من أجل الإسهام في اتخاذ القرارات بشأن الأفراد. ففي كثير من الحالات يمكن أن تكون الاختبارات حديثة التكوين أو التي لم تتطور إلى درجة ملائمة ذات فائدة كبيرة في التجارب. فليس من المهم أن يكون مؤلف الاختبار أو ناشره قد وضع له معايير جيدة، طالما أن الباحث لا يهتم بالمستوى العام لقدرة وتحصيل

المفحوصين، ولكنه يهتم فقط بما إذا كان مستوى المجموعة التجريبية أعلى من مستوى المجموعة الضابطة. كما أن موضوع الصدق قد لا يكون هاماً جداً إذا كان المجرب يخطط لاستخدام النتائج فقط كطريقة للتحقق من أحد الفروض النظرية. كما أنه ليس من المهم أن يكون للاختبار معامل ثبات عالٍ إذا كان المطلوب فقط هو مجرد مقارنة متوسطات بعض الجماعات، فأخطاء الصدفة سوف تلغي بعضها بعضاً، بحيث تبقى درجة متوسطة دقيقة إلى حد ما حتى ولو لم تكن الدرجات الفردية دقيقة.

وبالإضافة إلى القصور الذي يفرض نفسه على استخدام الأرقام التي لا تكوّن مقياساً للنسبة، فهناك مشكلة أخرى معقدة على وجه خاص يجب أن نتذكرها دائمًا عندما نستخدم الاختبارات في البحوث التجريبية. هذه المشكلة هي أنه من الصعب، إن لم يكن من المستحيل، قياس التغير بدقة بواسطة الاختبارات النفسية. فليست هناك مشكلة إذا كان موضوع البحث من النوع البسيط الذي يجاب عليه بالنفي أو الإيجاب، مثل «هل التدريب على التمييز الإدراكي يؤدي إلى زيادة واضحة في نسبة الذكاء؟» ولكن الباحثين يجدون مشاكل في التجارب التي تصمم لمقارنة التغير الناتج عن معالجة معينة بالتغير الناتج عن معالجة أخرى. وأحد مصادر هذه الصعوبات ما نسميه تأثير السقف. فمعظم الاختبارات مؤلفة من أسئلة يجيب عليها المفحوص، والأفراد الذين يعرفون معظم الإجابات الصحيحة من قبل يصبح من الصعب عليهم إحراز تقدم واضح مهما كان التغيير الذي أحدثه الموقف التجريبي فيهم. ومصدر آخر للمشكلة هو موضوع عدم الثبات. فإذا كانت الدرجة الإبتدائية التي يحصل عليها كل مفحوص تحت ظروف التجريب غير دقيقة بسبب أخطاء الصدفة (جميع درجات الاختبارات غير دقيقة، إلى حد ما، كما سبق أن بيّنا ذلك)، والدرجة النهائية للمفحوص أيضاً غير دقيقة، فإن طرح إحداهما من الأخرى يعطي درجة ذات قدر مضاعف من عدم الثبات. فالفرد (س)، على سبيل المثال، قد يحصل على درجة أعلى مما يجب في المرة الأولى، وعلى أقل مما يجب في المرة الثانية، فإن هذا يجعل التغير الذي أحدثته التجربة يبدو تافهاً، أو حتى يبدو في الاتجاه المخالف في هذه الحالة. والفرد (ص)، من جهة أخرى، لم يحالفه الحظ في المرة الأولى، ولكن حالفه الحظ في المرة الثانية، فهذا يؤدي الى ما يبدو كأنه زيادة إيجابية كبيرة في هذه الحالة. وهنا قد يستنتج الباحث غير المنتبه أن الفرد

(ص) تعلم أكثر من الفرد (س)، حتى ولو لم يكن هناك فرق في الحقيقة.

وقد اقترحت طرق متعددة لتفادي الأخطاء التي تؤدي إليها درجات التغير هذه. وأبسط هذه الطرق وأحسنها هي أن تصمم التجربة بطريقة لا يصبح من الضروري فيها استخدام درجات التغير على الإطلاق. فإذا كانت هناك عدة أنواع من المعالجات التجريبية يجب أن تتناولها التجربة، فالأفراد الذين يمثلون العينة الكلية يمكن تصنيفهم عشوائياً في مجموعات أ، ب، جـ، وهكذا. ثم نقارن بين متوسطات درجات هذه المجموعات على الاختبار في نهاية التجربة، وهذا سوف يجيب مباشرة على السؤال الذي يحاول البحث الإجابة عليه، حيث أنه لا يوجد هناك سبب يجعلنا نعتقد أنه توجد فروق منتظمة بين المجموعات منذ البداية.

تطبيقات الاختبارات والمقاييس ـ كلمة أخيرة

إن التقدير الكمي للسمات الإنسانية، وابتكار الطرق الإحصائية المناسبة لأنواع الأرقام التي نحصل عليها، وبناء الاختبارات لقياس القدرات وخصائص الشخصية، كل هذا يتطلب الإبداع، والتفكير الواضح، والشك العلمي الحذر. وهذه الصفات يحتاجها أيضاً من يريد أن يستخدم الاختبارات والمقاييس استخداماً مفيداً. فالاختبارات والمقاييس أدوات إنسانية صممت من أجل أغراض إنسانية. وهي بمفردها لا تحسم حواراً نظرياً، ولا تعالج مريضاً، ولا تعلم أطفالاً، ولا تحل مشكلات اجتماعية. ولكنها في أيدي الإخصائيين الماهرين الذين يفهمونها، تستطيع أن تساعدنا في جميع هذه المهام. فكل إنسان ـ رجلاً كان أو امرأة ـ عليه أن يتخذ قرارات حول مسائل عديدة أثناء تأدية عمله واتصاله بالناس الآخرين. كما أن هناك العديد من المشكلات الاجتماعية تبحث عن حلول. ومدى سلامة هذه القرارات أو هذه الحلول سوف يعتمد إلى حد كبير على دقة وإبداع صانعي القرارات والباحثين، هؤلاء الذين يقيمون الناس، والذين يخططون وينفذون برامج البحوث الأساسية. وإن ما تسهم به الاختبارات والمقاييس في هذه الأعمال الهامة هو الذي يبرر وجود هذه الاختبارات والمقاييس.

Selected Readings

ANASTASI, ANNE, ed. *Testing Problems in Perspective.* Washington: American Council on Education, 1966.

In 1936 the American Council on Education inaugurated a series of annual invitational conferences on testing problems. This book is a 25th anniversary volume that brings together many of the major issues and ideas participants discussed.

BUROS, E. K. JR., ed. *Mental Measurements Yearbook.* Highland Park, N. J.: Gryphon Press.

There have been six editions of this indispensable reference book in the years from 1941 to 1965. In each of them, essential information about each particular test is summarized, and two or more reviewers comment on its strengths and weaknesses.

CRONBACH, LEE J. *Essentials of Psychological Testing.* 3rd ed. New York: Harper & Row, Publishers, 1969.

This is one of the best complete surveys of the whole field of psychological testing. Additional information about any of the tests mentioned in the present book is presented, and a great many other tests are discussed in some detail.

GUILFORD, J. P. *Intelligence, Creativity and their Educational Implications.* San Diego, Calif.: Robert R. Knapp, 1968.

Guilford has been one of the most outstanding contributors to measure-

ment theory and technique. This book will introduce the student to some of his ideas.

HAMMOND, KENNETH R., and JAMES E. HOUSEHOLDER. *Introduction to the Statistical Method.* New York: Alfred A. Knopf, Inc., 1962.

In this book a psychologist and mathematician have combined their efforts to present the basic ideas of statistics in an exceptionally lucid and interesting way.

HUNT, J McV. *Intelligence and Experience.* New York: The Ronald Press Company, 1961.

To date, this is the best synthesis of a large amount of research designed to show what intelligence is and how it grows.

PETERSON, JOSEPH. *Early Conceptions and Tests of Intelligence.* Yonkers: World Book, 1925.

The early history of intelligence testing is covered here in much more detail than in other texts. It is often interesting for the student to understand the thinking of the people who began the work in a major area.

TYLER, LEONA E. *The Psychology of Human Differences.* New York: Appleton-Century-Crofts, 1965.

In this more extensive book, many of the ideas discussed briefly in the present volume are gone into in more detail, with research evidence.

المصطلحات الواردة بالكتاب
حسب ترتيب ورودها [1]

الفصل الأول

المصطلح	الصفحة	الأصل باللغة الإنجليزية
تجريبي	(١٧)	Emperical
يقدّر كميًا	(١٧)	Quantify
قياس ، مقياس	(١٧ ، ٢٢)	Measurement
احتمال	(١٩)	Probability
انفصام	(٢١)	Schizophrenia
استجابة القلق	(٢١)	Anxiety reaction
نسبة الذكاء	(٢٢)	Intelligence quotient (IQ)
مقياس التصنيف	(٢٣)	Nominal scale
الفئات	(٢٣)	Categories
مقياس الترتيب	(٢٣)	Ordinal scale

(١) يذكر المصطلح فقط عند أول وروده بالكتاب

المصطلح	الصفحة	الأصل باللغة الإنجليزية
الدرجات المئينية	(٢٣)	Percentile scores
مقياس الوحدات المتساوية	(٢٤)	Interval scale
مقياس النسبة	(٢٦)	Ratio scale

الفصل الثاني

المصطلح	الصفحة	الأصل باللغة الإنجليزية
اختلاف	(٢٩)	Variation
توزيع تكراري	(٢٩)	Frequency distribution
متوسط	(٣٠)	Average
نزعة مركزية	(٣٠)	Central tendency
منحنى الاحتمالات الاعتدالي	(٣١)	Normal probability curve
منحنى جاوس	(٣١)	Gaussian curve
إعتدالي (في الإحصاء)	(٣١)	Normal
ملتوٍ	(٣١)	Skewed
مدبب	(٣١)	Peaked
مفرطح	(٣١)	Stretched out
وسيط	(٣٣)	Median
دليل	(٣٣)	Index
منوال	(٣٣)	Mode
تشتت	(٣٤)	Variability
مدى	(٣٤)	Range
الانحراف المعياري	(٣٥)	Standard Deviation
التباين	(٣٦)	Variance
معامل الارتباط	(٣٦)	Correlatio Coefficient
معامل حاصل ضرب العزوم	(٣٦)	Product-Moment Coefficient
استدلالات	(٤٠)	Inferences
عينة	(٤٠)	Sample
مجتمع	(٤٠)	Population
تجريبي	(٤٠)	Emperical
توزيع المعاينة	(٤١)	Sampling Distribution
مقياس إحصائي ـ إحصاء	(٤٢)	Statistic

المصطلح	الصفحة	الأصل باللغة الإنجليزية
الخطأ المعياري	(٤٢)	Standard Error
دلالة إحصائية	(٤٥)	Statistical Significance
الإحصاء الاستدلالي	(٤٥)	Inferential Statistics

الفصل الثالث

المصطلح	الصفحة	الأصل باللغة الإنجليزية
علم النفس الفيزيائي	(٤٧)	Psychophysics
عتبة	(٤٨)	Threshold
اختبار	(٤٨)	Test
صدق (الاختبار)	(٥٠)	Validity
التفكير الاستدلالي	(٥١)	Reasoning
الاستعداد الميكانيكي	(٥١)	Mechanical Aptitude
المرونة المعرفية	(٥١)	Cognitive Flexibility
النضج الانفعالي	(٥١)	Emotional Maturity
المحك	(٥٢)	Criterion
معاملات الصدق	(٥٢)	Validity Coefficients
ثبات (الأختبار)	(٥٦)	Reliability
مهارة الأصابع	(٥٧)	Finger Dexterity
اختبارات مقننة	(٦٠)	Standardized Tests
درجات مشتقة أو محوّلة	(٦٠)	Derived or Transformed Scores
معايير	(٦٠)	Norms
استعدادات	(٦٠)	Aptitudes
تحصيل	(٦٠)	Achievement
معايير مئينية	(٦٠)	Percentile Norms
درجة خام	(٦٠)	Raw Score
درجة معيارية	(٦٠)	Standard Score
اختبار خالي من العوامل الحضارية	(٦٤)	Culture-free Test
مقياس المحك	(٦٤)	Criterion Measure
الصدق التنبئي	(٦٤)	Predictive Validity
التحليل العاملي	(٦٥)	Factor Analysis

الفصل الرابع

المصطلح	الصفحة	الأصل باللغة الإنجليزية
فطري	(٦٧)	Innate
طلاقة	(٦٩)	Fluency
النمو العقلي	(٧٠)	Mental Growth
تذكر مباشر	(٧٠)	Immediate Memory
معايير العمر	(٧٢)	Age Norms
العمر العقلي	(٧٢)	Mental Age
ثبات نسبة الذكاء	(٧٢)	Constancy of the IQ
راشد	(٧٣)	Adult
اختبار فرعي	(٧٥)	Subtest
اختبارات عملية	(٧٥)	Performance Tests
اختبارات لفظية	(٧٥)	Verbal Tests
نموذج الدرجات	(٧٦)	Score Pattern
بروفيل الدرجات	(٧٦)	Score Profile
قلق	(٧٧)	Anxiety
فروض	(٧٧)	Hypotheses
تنبؤ	(٧٧)	Prediction
موهبة	(٧٨)	Talent
اختبار جمعي	(٧٨)	Group Test
اختبار فردي	(٧٨)	Individual Test
بيت محتضن (أو بديل)	(٨٠)	Foster home
عمر نمائي	(٨١)	Developmental age
عمر زمني	(٨١)	Chronological age
خاصة (أو صفة)	(٨٢)	Quality
إمكانية	(٨٢)	Potential
القدرة المكانية	(٨٢)	Spatial
الإبداع	(٨٢)	Creativity
معايير	(٨٣)	Norms

المصطلح	الصفحة	الأصل باللغة الإنجليزية
التحصيل	(٨٥)	Achievement
درجة النغم (أو طبقة الصوت)	(٨٥)	Pitch
إيقاع	(٨٥)	Rhythm
اختبارات حرف	(٨٦)	Trade tests
القائم بمقابلة التوظيف	(٨٦)	Employment interviewer
العلاقات المكانية	(٨٨)	Spatial Relations
مهارة (وعلى الأخص المهارة اليدوية)	(٨٨)	Dexterity
التآزر الحركي	(٨٨)	Motor coordination
الفهم الميكانيكي	(٨٩)	Mechanical comprehension
تقدير	(٨٩)	Rating
درجة تُساعية	(٨٩)	Stanine
صدق المحتوى (أو المضمون)	(٨٩)	Content validity
التوجيه المهني	(٩١)	Vocational guidance
اهتمامات	(٩١)	Interests
التحليل العاملي	(٩٣)	Factor analysis
عامل	(٩٤)	Factor
تحليل العمل	(٩٥)	Job analysis
الذكاء العام	(٩٦)	General intelligence
القدرة اللغوية	(٩٦)	Verbal ability
القدرة العددية	(٩٦)	Numerical ability
القدرة المكانية	(٩٦)	Spatial ability
إدراك الشكل	(٩٦)	Form perception
الإدراك الكتابي	(٩٦)	Clerical perception
التآزر الحركي	(٩٦)	Motor coordination
مهارة الأصابع	(٩٦)	Finger dexterity
المهارة اليدوية	(٩٦)	Manual dexterity
درجات دنيا	(٩٦)	Minimum scores

المصطلح	الصفحة	الأصل باللغة الإنجليزية
نماذج القدرات المهنية	(٩٦)	Occupational ability patterns
تمييز	(٩٩)	Discrimination
تحليل البنود	(١٠٠)	Item analysis
اختبار متعدد الاختيار	(١٠٠)	Multiple-choice test
خطأ معياري	(١٠٠)	Standard error
الصدق الظاهري	(١٠١)	Face validity
التفكير اللغوي	(١٠١)	Verbal reasoning
التفكير التجريدي	(١٠١)	Abstract reasoning
التفكير المكاني	(١٠١)	Space reasoning
التفكير الميكانيكي	(١٠١)	Mechanical reasoning
السرعة والدقة الكتابية	(١٠١)	Clerical speed and accuracy
جماعة معيارية (التي اشتقت على أساسها المعايير)	(١٠٢)	Norm group
قدرة خاصة	(١٠٢)	Special ability

الفصل السادس

المصطلح	الصفحة	الأصل باللغة الإنجليزية
ضبط النفس	(١٠٦)	Self-control
إثارة	(١٠٦)	Irritation
إحباط	(١٠٦)	Frustration
استخبار	(١٠٦)	Questionnaire
استبيان	(١٠٦)	Inventory
الثبات الانفعالي	(١٠٧)	Emotional stability
توافق	(١٠٧)	Adjustment
العصابية	(١٠٧)	Neuroticism
اتجاه عقلي للاستجابة	(١٠٨)	Response set
الرغبة في القبول الاجتماعي	(١٠٩)	Social desirability
المسايرة	(١٠٩)	Acquiescense
الاعتماد على الغير	(١١٠)	Dependability
أثر (أو تأثير) الهالة	(١١٠)	Halo effect
مقياس تقدير	(١١٠)	Rating scale

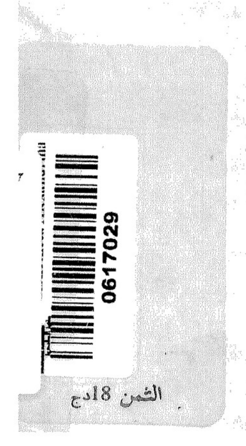
المصطلح	الصفحة	الأصل باللغة الإنجليزية
اتصال ـ تدرج متصل	(١١٠)	Continuum
بارانوي	(١١١)	Paranoid
نفسجسمي ـ سيكوسوماتي	(١١١)	Psychosomatic
دافع الإنجاز	(١١١)	Achievement motive
حاجة	(١١٢)	Need
انتماء	(١١٢)	Affiliation
بارانويا	(١١٢)	Paranoia
تجريبي	(١١٢)	Empirical
خبرة	(١١٢)	Experience
انطواء	(١١٤)	Introversion
انبساط	(١١٤)	Extraversion
توازن ، اتزان	(١١٧)	Poise
سيطرة ، تسلط	(١١٧)	Ascendancy
تأكيد الذات	(١١٧)	Self-assurance
سيطرة	(١١٧)	Dominance
مكانة	(١١٧)	Status
مقدرة اجتماعية	(١١٧)	Sociability
حضور اجتماعي	(١١٧)	Social presence
تقبل الذات	(١١٧)	Self-acceptance
التطبيع الاجتماعي	(١١٧)	Socialization
قيم	(١١٧)	Values
ضبط النفس	(١١٧)	Self-control
تحمل	(١١٧)	Tolerance
مسايرة موافقة	(١١٧)	Conformance
مرونة	(١١٧)	Flexibility
الأساليب الإسقاطية	(١١٩)	Projective techniques
بناء ، تكوين	(١٢٠)	Structure
عياني	(١٢٠)	Concrete
توحّد	(١٢٢)	Identification

الفصل السابع

المصطلح	الصفحة	الأصل باللغة الإنجليزية
الإرشاد النفسي	(١٢٦)	Counseling
اختيار (أو انتقاء)	(١٢٨)	Selection
جدول توقع	(١٢٩)	Expectancy lable
تصنيف	(١٣٠)	Classification
فصام (سيزوفرينيا)	(١٣١)	Schizophrenia
ذهان عضوي	(١٣١)	Organic psychosis
الرسم الكهربائي للمخ	(١٣١)	Electroencephalogram
تهتهة ، لجلجة	(١٣١)	Stutter, stuttering
علم النفس الفارق	(١٣٤)	Differential psychology
توائم أحادية اللاقحة	(١٣٦)	Monozygotic twins
توائم متماثلة	(١٣٦)	Identical twins
مورثات ، جينات	(١٣٦)	Genes
توائم أخوية	(١٣٦)	Fraternal twins
توائم ثنائية اللاقحة	(١٣٦)	Dizygotic twins
دراسة الحالة	(١٣٩)	Case study
دراسة ارتباطية	(١٤٠)	Correlational study
متغيرات مستقلة	(١٤١)	Independent variables
مليثانية : جزء من ألف من الثانية	(١٤١)	Millisecond
متغيرات تابعة	(١٤١)	Dependent variables
ضغط (شدّة)	(١٤١)	Stress
متغيرات ضابطة	(١٤١)	Control variables
تأثير السقف	(١٤٢)	Ceiling effect
تقدير كمي	(١٤٣)	Quantification

أنجز طبعه على مطابع
ديوان المطبوعات الجامعية
الساحة المركزية بن عكنون ــ الجزائر

رقم النشر: 1978 —— 05 —— 85


06617029
الثمن 18دج